(2)

THE APPLICATION OF RESPONSE SURFACE METHODOLOGY TO THE OPTIMIZING OF THE REACTIVE ION ETCHING OF POLYSILICON

APPROVED:

Supervisor: Al F. Tasch

Al F. Tasch

Willis A. Adcock

Willis A. Adcock

August 1990 Thesis

The Application of Response Surface Methodology to the Optimizing of the Reactive Ion Etching of Polysilicon

CPT John A. Stine

Performed by John A. Stine while a graduate student at the University of Texas-Austin

U.S. Army Student Detachment
FT Benjamin Harrison, IN 46216-5820

Approved for public release; distribution is unlimited A

The goal of this research was to learn an empirical procedure for process optimizing and to apply it to the reactive ion etching (RIE) of polysilicon. The procedure chosen was Response Surface Methodology (RSM). RSM is explained in terms of the techniques of experimental design; factor evaluation using designed experiments; model building using experimentally obtained data and least squares regression; and model analysis using the method of steepest ascent, canonical analysis, and the method of ridge analysis. These techniques are combined into a generic flow chart that outlines the organization of a response surface study for process optimizing. This organization is then used to plan and begin the optimizing of the RIE of polysilcon in a chlorine and helium plasma. Two responses were critical to the optimizing of this process, polysilicon-to-oxide selectivity and anisotropy. The four factors: helium flow rate, chlorine flow rate, electrode power, and chamber pressure, were found to be significant in their effects on these responses and were used as the variable parameters for optimizing the process. But the study revealed that a fifth parameter, the initial photoresist profile, had a dominant effect on the anisotropy results due to the low polysilicon-to-photoresist selectivity. The project, therefore, fell short of optimizing the process. However, suggestions are made on how to limit the influence of this biasing element and recommendations are given on how to continue the optimizing study. The study further reveals that there is a region in the parameter space in which both responses can be optimized simultaneously.

Response Surface Methodology, Empirical model building, Process Optimizing, Least Squares Regression, Design of Experiments, Reactive Ion Etching 246

UNCLASSIFIED UNCLASSIFIED UNCLASSIFIED SAR

DEDICATION

To my wife Martha and my sons, Andrew, Harry, and Brian.

THE APPLICATION OF RESPONSE SURFACE METHODOLOGY TO THE OPTIMIZING OF THE REACTIVE ION ETCHING OF POLYSILICON

by

JOHN ANDREW STINE B.S., M.S.

THESIS

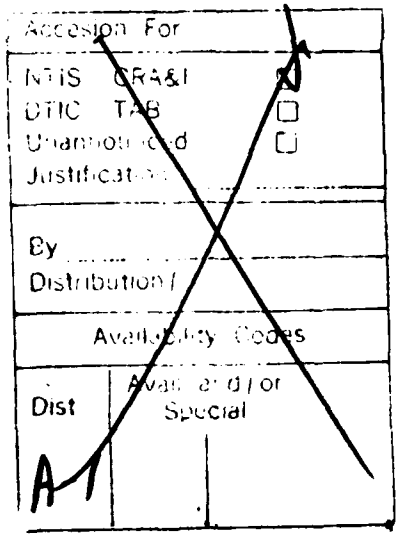
Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of
Master of Science in Engineering

THE UNIVERSITY OF TEXAS AT AUSTIN

August, 1990

ACKNOWLEDGEMENT

I wish to thank my supervisor, Dr. Al F. Tasch, for the quidance and assitance he provided throughout the course of this work. I also wish to thank a fellow student, GregYeric, for his assistance in training me to use the equipment in the lab.



Date submitted to Committee: May 19, 1990

ABSTRACT

THE APPLICATION OF RESPONSE SURFACE METHODOLOGY TO THE OPTIMIZING OF THE REACTIVE ION ETCHING OF POLYSILICON

by

JOHN ANDREW STINE B.S., M.S.

SUPERVISING PROFESSOR: AL F. TASCH

The goal of this research was to learn an empirical procedure for process optimizing and to apply it to the reactive ion etching (RIE) of polysilicon. The procedure chosen was Response Surface Methodology (RSM). RSM is explained in terms of the techniques of experimental design; factor evaluation using designed experiments; model building using experimentally obtained data and least squares regression; and model analysis using the method of steepest ascent, canonical analysis, and the method of ridge analysis. These techniques are combined into a generic flow chart that outlines the organization of a response surface study for process optimizing. This organization is then used to plan and begin the optimizing of the RIE of polysilcon in a chlorine and helium plasma. Two responses were critical to the optimizing of this process, polysilicon-to-oxide selectivity and anisotropy. The four factors: helium flow rate, chlorine flow rate, electrode power, and chamber pressure, were found to be significant in their effects on these responses and were used as the variable parameters for optimizing the

process. However, the study revealed that a fifth parameter, the initial photoresist profile, had a dominant effect on the anisotropy results due to the low polysilicon-to-photoresist selectivity. As a result, the optimum process parameters were not identified. However, suggestions are made on how to limit the influence of this biasing element, and recommendations are given on how to continue the optimizing study. The study further reveals that there is a region in the parameter space in which both responses can be optimized simultaneously.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Ever since Robert Noyce and Jack Kilby invented the integrated circuit in 1958, scientists and engineers worldwide have relentlessly pursued putting more and more transistors onto small semiconductor chips. The fruit of their efforts has been the logarithmic increase in the density of transistors on the chips over the past 30 years. But now, as the dimensions of the transistors have shrunk below the one micron level and as the global competition has sharply increased in electronics manufacturing, the issues in research are not only to increase the density of the transistors on the chips but also to improve the US semiconductor industry's ability to manufacture them.

Among the efforts to improve the manufacturability of integrated circuits are studies directed toward the development and improvement of new and old technologies. Inherent to these studies is the development of models that represent the processes of the technologies. With an accurate model, scientist and engineers can predict the responses of a process for different initial and operating conditions. In turn, they are able to predict the set of conditions that will allow the process to have the optimum response. The best models for this purpose are physically based. The development of physical models, however, requires an intimate understanding of the physical mechanisms involved in the process. This knowledge is not always available and may be difficult to determine due to the multitude of reactions and interactions in a process. Therefore, an alternative approach is used to understand such processes. The approach is empirically based; models are developed directly from experimental data. The research of this thesis is an

example of the use of these empirical techniques to both understand and optimize processes.

The goal of the research discussed in this thesis has been to learn the empirical techniques used to understand and optimize processes and to apply them to a semiconductor fabrication process. The experimental, mathematical, and statistical methods used in process optimization have long been a neglected topic in US universities and the US semiconductor industry. Only now are both industry and research institutions emphasizing that these skills be included in undergraduate and graduate curricula. Since this thesis research is one of the initial projects using these methods at the University of Texas - Austin, it includes not only the results of the optimization effort but also attempts to present the methodology. It is the intent of the thesis to explain not only the research results and conclusions, but also to serve as a guide to future researchers involved in similar projects. Therefore, the thesis has two parts: Chapters 1, 5, 6, and 7 deal specifically with the research; and, Chapters 2, 3, and 4, discuss the methodology of process optimization. Each part could stand alone, but together they provide both the theory and a practical example of understanding and optimizing processes using empirical techniques.

Chapter 1 presents the issues of polysilicon etching and describes the process that was optimized in this thesis research. Chapters 2, 3, and 4 discuss the experimental, mathematical, and statistical tools used in the research. Chapter 5 provides a description of the overall project. It includes information on the order and precedence of all the activities of the research, as well as descriptions of how certain important activities were conducted,

such as those used to measure etch selectivity and anisotropy. Chapters 6 and 7 present the results of the experiments, the subsequent analysis, and the final conclusions.

The techniques used to optimize a process are neither hard to understand nor difficult to apply, but do involve considerable time and effort. Accordingly, a well thought-out plan is essential to minimizing the effort of the experimenter. Hopefully, this thesis will motivate and help future researchers in their efforts to understand process optimization, and serve as a guide for them to planning their research.

CHAPTER 1

POLYSILICON ETCHING

In Very Large Scale Integration (VLSI) and Ultra Large Scale Integration (ULSI) circuits, millions of miniature devices are integrated on single silicon chips. The ability to fabricate these chips is largely dependent upon the ability to define and control the effects of each process used to make them.[1] One process, polysilicon etching, is especially critical. The results of the polysilicon etch will ultimately affect how well the circuit performs or even whether the circuit works at all. The goal of this research was to optimize polysilicon etching in a reactive ion etching (RIE) process using a chlorine and helium plasma. The process is specifically well suited for etching polysilicon; however, the factors that make it such are not well understood. For this reason, experimental methods were needed to study and optimize the process.

ETCHING PROPERTIES

Three properties of importance when comparing etching processes are etch rate, selectivity, and isotropy. Etch rate is simply the time rate at which a layer is removed during etching. Selectivity is defined as the ratio of

1 Throughout this thesis four words are used, which, in some cases, could be considered synonymous. These words are effects, responses, parameters, and factors. To help the reader avoid confusion, a single convention of using these words is used throughout the thesis. Parameters and factors are synonymous and refer to the conditions under which a process operates. Effects and responses refer to the results of a process. A response is a discrete measure of the result of a run of a process. An effect is a measure of how a response changes from run to run when parameters are changed.

etch rates of different materials. A selective process has a high etch rate for the etched layer and a low etch rate for the other layers that are exposed during the process. Isotropy is defined as the directionality of the etch. A process which etches in all directions at the same rate is considered isotropic, while a process which etches in a preferred direction at a higher rate than other directions is considered anisotropic.

An example of an isotropic etching process is wet etching. In wet etching, the silicon wafer is submerged in an etching solution. The etching occurs in three steps: 1) a reactant is transported to the etched surface; 2) a chemical reaction occurs; and then 3) the reaction products move away from the surface. This process is isotropic, since the reaction is independent of direction. It will occur normal to the exposed surface regardless of the direction of that normal. (Some etching processes are anisotropic. They etch faster in a preferred crystallographic direction of a single crystalline structure. Obviously, no such wet etch process exists for polycrystalline silicon.) An advantage of wet etching is its selectivity. The solutions used for the etch can be mixed such that a chemical reaction only occurs preferentially with the layer desired.

An example of an anisotropic etch is ion beam milling. In ion beam milling, ions are accelerated into the surface being etched. When the ions strike the surface, they eject atoms from the surface similar to a racked set of pool balls struck by a queue ball. This billiard-like scattering is called sputtering. The rate of the etching process depends on the direction and energy of the striking ions. Since the direction is easily controlled, it is

anisotropic. But, since it relies on a physical reaction, such an etching process is not very selective.

POLYSILICON ETCHING

Polysilicon etching requires a process that is both selective and anisotropic. Polysilicon is used as the gate and interconnect material for both CMOS and NMOS circuits. In the fabrication of these circuits, the polysilicon also serves as the "self aligned" mask for the subsequent implantation of the impurities that form the source and the drain of the MOS field effect transistors (FET). The polysilicon, therefore, defines the dimensions of the gate. These dimensions are very critical, since they determine both the speed and the logic levels of the circuit. Figure 1.1 illustrates the desired results of

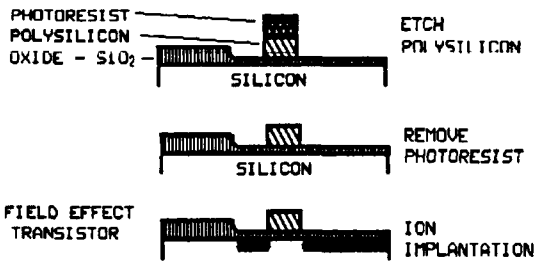


Figure 1.1 An ideal polysilicon etch and gate masking

the polysilicon etch. The polysilicon is etched anisotropically, retaining the width of the photoresist. In the subsequent ion implantation, the source and drain are separated by a gate the width of the polysilicon. Figure 1.2 shows the results of wet etching. There is undercutting of the polysilicon beneath the

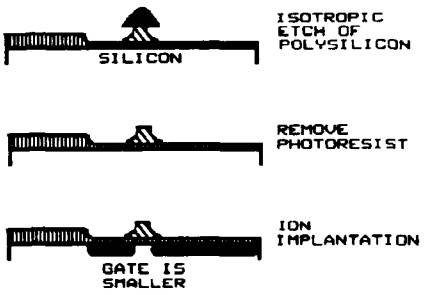


Figure 1.2 The wet etch of polysilicon and its effect on doping

photoresist. In the subsequent ion implantation, some of the impurity atoms penetrate the polysilicon, resulting in a narrower gate. Figure 1.3 shows the results of ion beam milling. This figure illustrates two undesirable effects. If

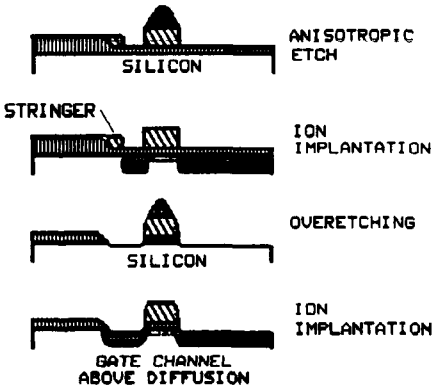


Figure 1.3 The ion beam milling of polysilicon and its effect on doping

the etch is continued only to the point where mass spectroscopy indicates that the underlying SiO_2 is exposed, stringers may still exist. Stringers are residues of conformal films that occur at steps on the wafer surface after etching. The vertical thickness of conformal films at steps is greater than over

planar areas. Since an anisotropic etch only etches in the vertical direction, the etch must be extended in time to account for this additional thickness. If the etch is not extended and stringers remain afterwards, several adverse effects may occur. First, polysilicon lines running parallel may be shorted. Second, the presence of polysilicon may mask the doping of some regions on the wafer. And third, the polysilicon may interfere with metal contacts or even short the metal interconnects with the polysilicon interconnects. If the etch is extended in a non-selective process, the second result, overetching, may occur. The underlying SiO_2 and some of the substrate silicon may be etched. If etched excessively, the depth of the source and drain relative to the gate dielectric-silicon interface may be increased to the point where the device performance is degraded. One or both of these problems would occur in ion beam milling. Since gate oxides are very thin, stopping the etch prior to penetrating the oxide would be difficult. In addition, even if we could stop the etch, stringers would almost certainly remain. It would be almost impossible to obtain a reasonable yield with ion beam milling.

SOLUTION

Reactive Ion Etching (RIE) offers a viable means of achieving an anisotropic etch that remains selective. It does this by combining the chemical and physical elements seen in the wet etching and ion beam milling processes. The etching occurs in a plasma. The components of the plasma are both the reactant species and the ionsaccelerated into the etched layer.

The etching process occurs in one of two ways. In the first, ions are accelerated at a rate that damages the surface but avoids sputtering. As in

the chemical process, the reactants move to the surface and a chemical reaction occurs; but, the reaction occurs at a much faster rate at the points damaged by the bombarding ions. In the second way, a chemical reaction occurs at the surface in which the reactant species is absorbed, forming a layer that is more easily sputtered than the original. In both cases, the ion bombardment enhances the rate of the chemical reaction. Since the direction of ion bombardment can be controlled, the process is anisotropic, and, since the etch is dominated by a chemical reaction, the etch remains somewhat selective.

PLASMA

The etching environment for RIE is a plasma. A plasma is defined as a partially ionized gas that contains positve and negative charges and some non-ionized gas particles. A plasma is created by injecting a gas between two oppositely charged electrodes. When the electrodes receive rf power, charged particles that exist in the gas begin to accelerate. These particles react with the gas molecules resulting in five different reactions:

1.**Excitation.** Electrons of the gas molecules are excited. In their subsequent deexcitation, a photon is released and a glow may be observed.

Example. $e^{-*} + He \rightarrow He^{*} + e^{-}$

$He^{*} \rightarrow He + h\upsilon$ [2]

2 * indicates that a particular species is in an excited energy state so that its energy is greater than that at thermal equilibrium.

2. **Ionization**. An energetic electron strikes a gas molecule, knocking another electron from the molecule and thus creating an ion.

Example. $e^{-*} + Cl_2 \rightarrow Cl_2^- + 2e^-$

3. **Dissociation**. An energetic electron strikes a gas molecule, separating it into two radicals.

Example $e^{-*} + Cl_2 \rightarrow Cl + Cl + e^-$

4. **Dissociative Ionization**. An energetic electron strikes a gas molecule, separating it into ions and radicals.

Example $e^{-*} + Cl_2 \rightarrow Cl^+ + Cl + 2e^-$

5. **Dissociative Attachment**. An electron strikes a gas molecule, separating it into ions. Example $e^{-*} + Cl_2 \rightarrow Cl^+ + Cl^- + e^-$

Two reactions that occur at a wafer surface cause etching.

1. **Atom Abstraction.** An atom of the wafer reacts with a component of the plasma to form a new molecule.

Example. $Si(s) + Cl_2 \rightarrow SiCl_2(g)$ or $SiCl_2(s)$

2. **Sputtering**. An ion strikes the wafer surface, ejecting atoms of the target.

Example. $He^{+*} + SiCl_2(s) \rightarrow SiCl_2(g) + He^+$

While etching is underway, the final plasma consists of the following: a) etch gas molecules (70-98% of the total species in the chamber); b) etch product molecules (2-20%); c) radicals (0.1-20%); and d) charged species, including positive ions, electrons , and negative ions (0.001 -0.01%). Of these, the radicals are responsible for most of the etching. So "reactive ion etching" may be more aptly called "ion assisted etching," since the ions themselves are not responsible for most of the etching; rather, they enhance the reaction rate of the radicals [6:544].

ENERGETIC ION ASSISTANCE

Two mechanisms exist by which energetic ions are believed to assist etching.

1. Relatively high energy ions, 50ev, damage the surface being etched. Etching occurs faster at these areas.
2. Lower energy ions provide energy to desorb nonvolatile polymer layers that deposit on the surface during plasma etching. These layers would protect against etching if not removed [6:533].

CONTROLLING RIE

Many different parameters can be changed to affect RIE. The goal is to determine the set of parameters that will achieve a selective and anisotropic etch. Selectivity is largely determined by the plasma chosen for the etching processes. Anisotropy is largely determined by the nature of the ion bombardment, which is most directly controlled by the power delivered to

the electrodes. Unfortunately, other factors affect these properties as well. The optimization of RIE involves the balancing of these factors to achieve the best possible combination of etch rate, selectivity, and anisotropy.

THE SPECIFIC PROCESS

The Reactive Ion Etcher in the UT Microelectronics Laboratory is a Plasma Technology RIE 80 M specifically configured to etch polysilicon using a chlorine and helium plasma. Figure 1.4 shows a diagram of its set-up.

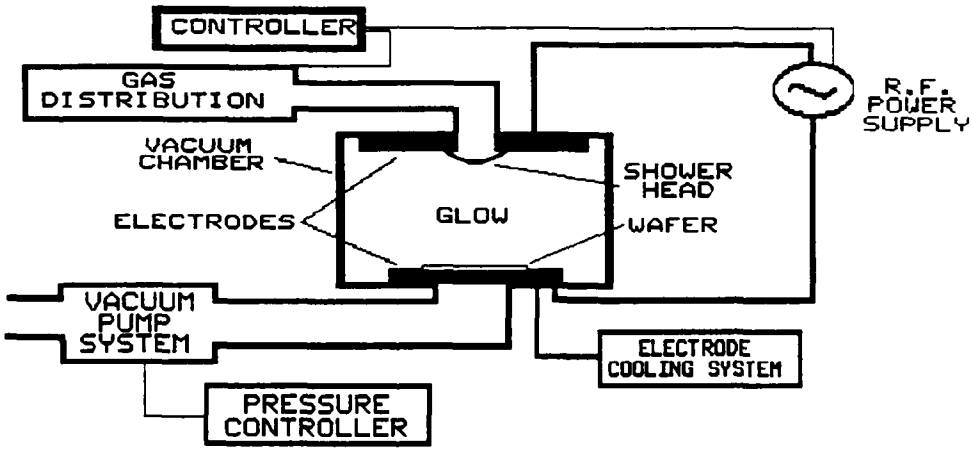


Figure 1.4 Plasma Technology RIE 80 M Reactive Ion Etcher

The choice of a chlorine and helium plasma was predetermined, although this equipment could surely be used with other gases. Industry has successfully used chlorine and helium etch recipes for a number of years. The great attraction of these recipes is their high selectivity to polysilicon over oxide. But, etching in a chlorine and helium plasma has its drawbacks. It is

not very selective to photoresist and, due to its selectivity over oxides, another process must be introduced to etch away any native or sacrificial oxides on the polysilicon. Thus, the etching process actually has two parts, one for removing any native or sacrificial oxides using a CF_4 plasma and one for etching the polysilicon using a He + Cl_2 plasma.

The process for etching polysilicon with the RIE 80 M system is as follows.

1. The wafer to be etched is placed on the bottom electrode in a recessed area with the surface to be etched facing up. This electrode is the powered electrode.
2. The chamber is evacuated and purged with nitrogen.

FIRST ETCH FOR OXIDES

3. The chamber is filled with CF_4 until a steady gas flow rate and a constant chamber pressure are achieved (set-up time). Then the electrodes are turned on for a predetermined time, and an etch ensues.
4. When the etching is finished, the chamber is evacuated.

SECOND ETCH FOR POLYSILICON

5.The chamber is filled with He and Cl_2 at predetermined proportions until the desired flow rate and chamber pressure are achieved (set-up time). Then the electrodes are turned on for a predetermined time and an etch ensues.
6. When the etching is finished, the chamber is evacuated and purged with nitrogen.
7. The wafer is removed.

Four parameters are controlled by the controller: the power to the electrodes, the gas flow rates, the set-up times, and the etch times. Two additional parameters, external to the etcher controller, that can also be controlled: the electrode coolant temperature and the chamber pressure. The etcher controller can control, separately, the set-up times, the etch times, the rf power, and the gas combinations of up to three etching processes. These can all vary between processes. The gas flow rates, chamber pressure, and coolant temperature can be varied between processes, but they can only be changed manually. (In this process, changing the gas flow rates between processes was not an issue, since none of the gases were used in both processes.) The following is a summary of the parameters that could be varied:

CF_4 ETCH	**He + CL_2 ETCH**
-set-up time	-set-up time
-electrode power	-electrode power
-chamber pressure*[3]	-chamber pressure*
-etch time	-etch time
-electrode coolant temperature*	-electrode coolant temperature*
-CF_4 flow rate	-He flow rate
-Cl flow rate	

3 The factors identified by the * must be changed manually between the CF_4 and the He + Cl_2 etches.

This list shows 13 different parameters that could affect the final polysilicon etch characteristics. Certainly, some of these are not very significant. So, the first step in studying this etch process was to reduce the list to a manageable size. The first step in optimizing a process is to evaluate the parameters. The evaluation involves three different actions: deleting any parameters that the experimenter believes are insignificant or impractical to vary; conducting experiments to verify the significance of parameters the experimenter is uncertain of; and, making a best guess as to what effects each significant parameter will contribute as it is varied. This last step is important. It reduces the size of the study and provides the experimenter the insight into the nature of the process which allows him to identify problems and critical phenomena. The following is the evaluation of the process that I conducted prior to beginning any experiments.

I decided to ignore all of process one, the CF_4 process, by making the assumption that it would have a minimal effect on the outcome of the polysilicon etch other than to vary the initial condition. Since I could keep all the parameters of the first process constant, I assumed I could also keep the initial conditions constant. (This assumption proved to be incorrect.)

I deleted the set-up time as a critical parameter, since it also only affects the initial condition. The set-up time is the period of time after the chamber evacuation and before the electrodes receive power when the etcher controller attempts to stabilize the gas flow rates and chamber pressure. This set-up time contributes nonlinearities to the etch as a result of two effects. In the first, etching occurs simply from the exposure of the wafer to the etch gases. Initially, since chlorine is very corrosive, I assumed that

there would be etching. However, I performed a quick experiment of exposing a wafer to the process set-up and found no evidence of etching. The second nonlinearity occurs during the brief period just as the electrodes receive power, when the plasma reactions have not reached the steady state condition. The nonuniform etching characteristics that occur during this period are unavoidable, but their variance can be minimized if the gas flow rates and the chamber pressure are stabilized prior to the charging of the electrodes. The minimum set-up time that allows all the gas flow rates and the chamber pressure to stabilize is preferred. The 41 second set-up times already set in the controllers proved to be adequate and were kept constant throughout the study.

I deleted the electrode coolant temperature as a critical parameter largely because of the difficulty involved in changing it. Changing the temperature involves an iterative process in which the adjustment of a control dial and the reading of the temperature is repeated until the desired temperatu e is achieved. Since this would have had to have been done between the CF_4 and the Cl_2 + He etching processes, I considered the adjustment very impractical. Besides, the purpose of the coolant is mostly to protect the electrodes from overheating rather than to specify the temperature at which the etch occurs.

I did not consider etch time to be a parameter of the process that could be varied for optimization. In actuality, etch time is a response and not a parameter. An optimum etch time exists, but it is a function of the etch rate and of the thickness of the etched layer.

This reduction left four parameters: the He flow rate, the Cl flow rate, power, and pressure. I immediately considered all of these to be significant. Below is my preliminary study of the expected effects of these four parameters.

HELIUM

Helium is an inert gas and, as such, does not chemically react with any of the exposed surfaces of a wafer. However, the presence of helium does affect the generation of ions and free radicals in the chlorine species, as well as provide a portion of the bombarding particles that strike the wafer surface. In the presence of enough chlorine, an increase in helium could increase both the polysilicon etch rate and the etch anisotropy. The mechanisms that increase the etch rate of polysilicon also increase the etch rates of the photoresist and the oxide layers. Therefore, I expected an increase in helium to decrease the selectivity but to increase the etch anisotropy.

CHLORINE

Chlorine is the gas that provides the radicals and ions responsible for the chemical portion of the etch. It is reasonable to assume that an increase in the chlorine concentration would increase etch rates and etch isotropy. Since the polysilicon is more sensitive to chlorine than to the oxide, the polysilicon/oxide selectivity is likely to increase.

POWER

The effect of power has long been understood in plasma etching. An increase in power will increase both the chemical and physical mechanisms of the etch, so that etch rates on all fronts will increase with increased power. The effects on anisotropy and selectivity are less obvious. Considering the power to be in the high range, I expected the increase in the physical mechanisms to exceed the increase in the chemical mechanisms. Accordingly, I expected the selectivity to decrease and the anisotropy to increase.

PRESSURE

The effects of pressure were the most difficult to predict. On the one hand, increased pressure would decrease both the physical and chemical mechanisms of the etch due to the expected decrease in the mean free path. But, studies have shown that increased pressure raises the absorption rate and the formation of $SiCL_2$ at the polysilicon surface [3: 23]. Since $SiCl_2$ is more easily sputtered from the wafer surface than silicon, I would expect a simultaneous increase in the polysilicon etch rate. So in the end, I expected an increase in pressure to increase the polysilicon selectivity over both oxide and photoresist and to decrease the etch anisotropy.

SUMMARY OF EXPECTED EFFECTS

	Polysilicon Etch Rate	Oxide Etch Rate	Photoresist Etch Rate	Polysilicon/ Oxide Selectivity	Polysilicon/ Photoresist Selectivity	Anisotropy
HELIUM	+	+	+	–	–	+
CHLORINE	+	0	+	+	0	–
POWER	+	+	+	–	–	+
PRESSURE	0	–	–	+	+	–

Table 1.1 The expected main effects of increasing parameters

After analyzing the process and the parameters, it was clear that there no simple way exists to predict an optimum operating point. The separate effects of each parameter, as well as the interactions between them, contribute to a plethora of reactions that would be very difficult to define. Additionally, certain uncontrollable factors that are equipment specific, such as how the gases flow in the chamber, or are etch specific, such as the pattern on the wafer to be etched, would make an exact study impractical. This situation was ideally suited for experimental methods of optimization.

Being uncertain how to optimize a process by experimental methods, I immediately attacked the problem by using a full 2^k factorial experiment with two replicates. This was a mistake. Experimental designs of this type are usually used to evaluate the effects and interactions of parameters to

determine if they are significant. Since I had already concluded that the individual effects and possibly some interactions of each of the parameters were significant, this course of action was wrong. Luckily, the effort was not wasted. To allow the reader to understand why and to follow the conclusions of this thesis, I now move from discussing my specific project to providing background on the experimental designs, the methods of empirical model building, and the analysis of experimental data that are combined into an overall methodology for optimizing processes.

CHAPTER 2

EXPERIMENTAL DESIGNS AND EMPIRICAL MODEL BUILDING

The experimental designs used in Response Surface Methodology (RSM) were largely developed by G. E. P. Box and his graduate students at the University of Wisconsin in the 1950's. The designs all involve the strategic selection of a set of experiments such that the nature of a response within their space can be modeled using the method of least squares. This chapter discusses three of these designs: full 2^k factorial, central composite, and 2^{k-p} fractional factorial. Although numerous other designs exist, this set is rather complete in its applicability to RSM and falls under the category of orthogonal designs, which are considered to be the most accurate. By applying the method of least squares, these designs allow the researcher to determine linear, interaction, or second-order models. In turn, these models allow the researcher to determine the significance of parameters to the modeled response, at which parameter settings the optimum response can be found, or in what way the parameters should be changed to find the optimum response. Chapters 3 and 4 cover these latter topics. For now, Chapter 2 concentrates on the basics, the design of experiments, the theory of the method of least squares, its application to model building, and, finally, the statistical analysis of a model's fit.

EXPERIMENT DESIGNS

The three designs discussed here are the 2^{k-p} fractional factorial, the full 2^k factorial, and the central composite. Each of these designs is an expansion of its predecessor, as listed. (The 2^{k-p} fractional factorial design is

a portion of the full 2^k factorial design, and the 2^k factorial design is a portion of the central composite design.) To explain them, however, it is necessary to describe the full 2^k factorial design and then to show how one can modify this design to form a central composite or a 2^{k-p} fractional factorial design.

In a full 2^k factorial design, all experiments are conducted about an arbitrary point defined by the values of each variable parameter of the process at that point. The design points at which the experiments are performed are located about this center point. One characteristic of a full 2^k factorial design is that each parameter of each experimental point may assume only one of two possible values, equidistant from the center value. The full 2^k factorial design comprises the set of experiments that includes all possible combinations of these parameters. As an illustration, consider a three- parameter process with parameters a, b, and c. The design center point is arbitrarily selected to occur at

$$a = 15,$$

$$b = 120,$$

$$c = 80,$$

and the values for the variation of each parameter are arbitrarily selected as

$$a = \pm 5,$$

$$b = \pm 30,$$

$$c = \pm 15.$$

Then the parameter set points for the experiment are combinations of

$$a = 15 \pm 5;$$

$$b = 120 \pm 30;$$

$$c = 80 \pm 15.$$

Since 2 levels for each of the three parameters exists, there are a total of 2^3 design points in the experiment,

(10, 90, 65),	(10, 90, 95),
(20, 90, 65),	(20, 90, 95),
(10, 150, 65),	(10, 150, 95),
(20, 150, 65),	(20, 150, 95),

and, thus, the name 2^k factorial experiment, k being the number of variable parameters.

Fundamental to an experimental design is the coding of the variable parameters.[1] In the full factorial design the parameter values are coded as +1 for the high points and as -1 for the low points using the equation

$$x_i = 2\ (p_i - \bar{p}_i) / d_i$$

where p_i is the parameter setting; $\bar{p}_i$ is the average of the two settings for a parameter (the center value); and d_i is the difference between the parameter settings. For example, for the parameter a

$$\bar{p}_1 = \bar{a} = (a^+ + a^-)/2 = (20 + 10) / 2 = 15$$

and

$$d_i = (a^+ - a^-) = (20 - 10) = 10\ .$$

1 This thesis uses the following convention for denoting the different aspects of each variable parameter setting. A lower case letter is a dimensional quantity that is the actual parameter setting. An upper case letter is a dimensionless quantity that refers to the high and low level of a parameter in an experiment and assumes the value of +1 or - 1. A numerically subscripted x, such as x_1, is a dimensionless quantity that can assume any value depending on the method used to encode the actual parameter setting. In encoding parameters, the normal practice is to have the encoded quantity x_1 equal +1 and - 1 at the high and low levels of the experimental design.

When a is high,

$$x_1 = 2\ (20 - 15)\ /\ 10 = +1,$$

and when a is low,

$$x_1 = 2\ (10 - 15)\ /\ 10 = -1\ .$$

With all variables coded, it is then possible to form the design matrix. For our three-parameter experimental design, the matrix is

$$D = \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \left[\begin{array}{ccc} -1 & -1 & -1 \\ +1 & -1 & -1 \\ -1 & +1 & -1 \\ +1 & +1 & -1 \\ -1 & -1 & +1 \\ +1 & -1 & +1 \\ -1 & +1 & +1 \\ +1 & +1 & +1 \end{array}\right] \end{array}.$$

From this matrix, we can form a second matrix, which we will use in our analysis, called the X matrix. We will discuss this matrix in more detail later, but, for purposes of illustration, the X matrix of the general linear model is written as

$$X = \begin{array}{c} \begin{array}{cccc} & x_1 & x_2 & x_3 \end{array} \\ \left[\begin{array}{c|ccc} 1 & -1 & -1 & -1 \\ 1 & +1 & -1 & -1 \\ 1 & -1 & +1 & -1 \\ 1 & +1 & +1 & -1 \\ 1 & -1 & -1 & +1 \\ 1 & +1 & -1 & +1 \\ 1 & -1 & +1 & +1 \\ 1 & +1 & +1 & +1 \end{array}\right] \end{array}.$$

Finally, Figure 2.1 shows a graphical representation of the 2^3 factorial design.

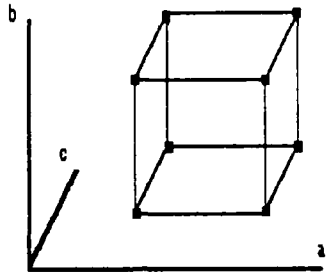


Figure 2.1 A Full 2^3 Factorial Design

In the central composite design, (2k + 1) additional experimental points are added to the full factorial design. One point is at the center of the experiment and the remainder are separated from the center by varying only one parameter by some constant value. (These points are on the axis of the design.) If we make the magnitude of our coded distance 2, the values for our a, b, and c parameters will be

$$|2| = 2(a - 15) / 10 \text{ so } a = 5, 25,$$

$$|2| = 2(b - 120) / 60 \text{ so } b = 60, 180,$$

$$|2| = 2(c - 80) / 30 \text{ so } c = 50, 110.$$

Figure 2.2 shows the illustration of the design.

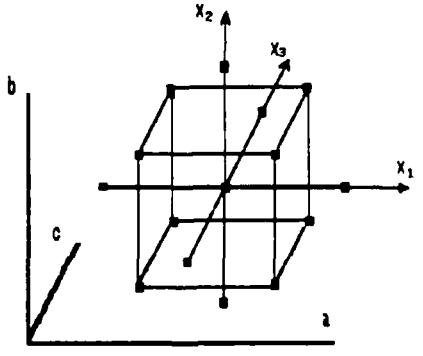


Figure 2.2 The Central Composite Design

which has the design matrix

$$D = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 \end{matrix} \\ \left[\begin{array}{ccc} +1 & -1 & -1 \\ +1 & -1 & -1 \\ -1 & +1 & -1 \\ +1 & +1 & -1 \\ -1 & -1 & +1 \\ +1 & -1 & +1 \\ -1 & +1 & +1 \\ +1 & +1 & +1 \\ +2 & 0 & 0 \\ -2 & 0 & 0 \\ 0 & +2 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & +2 \\ 0 & 0 & -2 \\ 0 & 0 & 0 \end{array}\right] \end{array} .$$

The fractional factorial design, as the name suggests, is a modified full factorial design with a reduced set of experimental points. The motivation for a fractional factorial design is simply this reduction in the size of the experiment. But, there is a trade-off; less information is available from the design than from the full factorial designs. The loss of information can be magnified if the fractional set of experiments is selected without care. So the main point in the description of the fractional factorial design is to explain the correct manner in which to choose the experiments.

There can be a reduction in the number of the full factorial design experiments by a factor of $(\frac{1}{2})^p$ where p is some integer less than k. The range of values that p may assume is dependent on the magnitude of k and on the ultimate purpose of the designed experiment. Clearly, a fractional factorial experiment used to build a first-order model may have a larger p than one used to build a second order model. (The latter model can only be built with the addition of the center and axial points of the central composite

design.) Therefore, statisticians have developed a method of classifying fractional factorial designs using a resolution number. But before explaining the method, it is important to understand how the fractional set of points is chosen from the set of points in a full factorial design. (The following explanation in not meant to fully explain why a particular set of points are used, but to illustrate the mechanics of choosing the points. It is assumed, as the reader reads later sections of the thesis, the reasons will become apparent.)

The experimental points of a fractional factorial design are selected by using one or more equations that relate the high and low settings of a parameter. An example of such an equation is $C = \pm AB$. This equation states that there is a full factorial design in the parameters A and B, and that the parameter C is set to the product of their encoded values. Let us walk through the setting-up of the design. First, the full factorial experiment in the parameters A and B is

POINTS	A	B
1	-1	-1
2	+1	-1
3	-1	+1
4	+1	+1

Now, expanding the design to the fractional factorial experiment in the three parameters A, B, and C using the relation $C = +AB$, we have

POINTS	A	B	C
1	- 1	- 1	+1
2	+1	- 1	- 1
3	- 1	+1	- 1
4	+1	+1	+ 1

.

This is the complete design. Since the design has only four points with three parameters, it is a 2^{3-1} fractional factorial design. The design matrix and graphical representation for the design are

$$D = \begin{matrix} x_1 & x_2 & x_3 \end{matrix} \begin{bmatrix} +1 & -1 & -1 \\ -1 & +1 & -1 \\ -1 & -1 & +1 \\ +1 & +1 & +1 \end{bmatrix} .$$

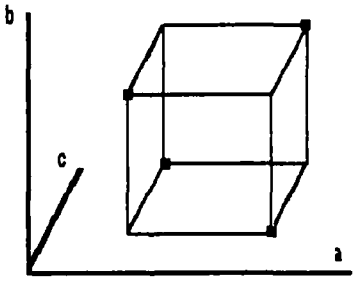


Figure 2.3 The 2^{3-1} Fractional Factorial Design

When an experimenter uses a fractional factorial design rather than a full factorial design, he loses the ability to make certain evaluations of the process he is analyzing. A replicated full factorial experiment allows an

experimenter to analyze the effects of each parameter in a process as well as the effects of all possible interactions of the parameters. When an experimenter uses a fractional factorial experiment, some of these effects will be indistinquishable. They are said to be confounded with each other. How the effects are confounded is an important issue in selecting a fractional factorial design. To determine confounding, one can use a simple table formed from the design points and the possible effects. The rows of the table correspond to the design points and the columns correspond to the effects. The table for our 2^{3-1} design is

POINTS	A	B	C	AB	AC	BC	ABC
1	- 1	- 1	+1	+1	- 1	- 1	+1
2	+1	- 1	- 1	- 1	- 1	+1	+1
3	- 1	+1	- 1	- 1	+1	- 1	+1
4	+1	+1	+1	+1	+1	+1	+1

The value in each cell of the table is the product of the parameters listed at the column heading at the settings of the experimental point defined by the row. For example, the entry in the cell beneath AB at experiment 2 is -1 since this is the product of the parameter settings A = +1 and B = - 1. The completed table defines the contribution of the response of each experiment to the final estimate of the effect. The contribution of the response of experiment point 2, y_2, to the estimate of the effect AB is $-y_2$. This will become more apparent later. When two effects are estimated with the same contributions from all the responses, they will be confounded. The table helps us identify the confounding. When two columns are identical, their effects are confounded. For example, the column beneath A is identical to the column

beneath BC, so A is confounded with BC. Similarly, B is confounded with AC, and C is confounded with AB. In short, the main effects are confounded with two-factor interactions.

The resolution of an experiment is a number (written in Roman numerals) given to define how the estimates of the effects are confounded. A design of resolution R is one in which the n-factor effect is confounded with no effect containing less than (R-n) factors. The above example is a resolution **III** experiment, since 1-factor effects are not confounded with any effects less than (3-1). In other words, the experiment does not confound the main effects with one another but does confound them with two-level interactions. Using this method of defining resolution makes it easy to define the nature of the confounding of other higher resolution experiments. For example:

1. A resolution **IV** experiment has no 1-factor effects that are confounded with any interaction less than (4-1) and has no 2-factor interactions that are confounded with any interaction less than (4-2). In other words, the main effects are not confounded with themselves or with any two-factor interactions, but the two-factor interactions are confounded with each other.

2. A resolution **V** experiment has no 1-factor effects that are confounded with any interaction less than (5-1) and has no 2-factor interactions that are confounded with any interaction less than (5-2). In other words, the main effects are not confounded with themselves, with any 2-factor interaction, or with any 3-factor interaction and the 2-factor interactions are not confounded with themselves.

Fractional factorial experiments may greatly reduce the amount of work one must do to evaluate a process, especially when working with a large number of parameters. If the intent is to ultimately fit a linear model to a process, a resolution **III** design is the smallest design that will yield reasonable results. The resolution **IV** design is the preferred design. If the intent is to fit a second-order model to the process, any design of resolution **V** or greater, combined with the center and axial points of the central composite design, may be used. Table 2.1 is a list of possible design generating equations for experiments with 3 to 7 parameters to obtain resolution **III**, **IV**, and **V** or higher experimental designs. Most texts on experimental design contain more complete tables.

Table 2.1 Design generators for fractional factorial experiments[2]

NUMBER OF PARAMETERS	FRACTION	RESOLUTION	NUMBER OF POINTS	EQUATIONS
3	2^{3-1}	III	4	C= ± AB
4	2^{4-1}	IV	8	D= ± ABC
5	2^{5-1}	V	16	E= ± ABCD
	2^{5-2}	III	8	D= ± AB E= ± AC
6	2^{6-1}	VI	32	F= ± ABCDE
	2^{6-2}	IV	16	E= ± ABC F= ± BCD
	2^{6-3}	III	8	D= ± AB E= ± AC F= ± BC
7	2^{7-1}	VII	64	G= ± ABCDEF
	2^{7-2}	IV	32	F= ± ABCD G= ± ABDE
	2^{7-3}	IV	16	E= ± ABC F= ± BCD G= ± ACD
	2^{7-4}	III	8	D= ± AB E= ± AC F= ± BC G= ± ABC

2 This table is a truncated and modified version of Table 11-8 found in [4] on page 338.

When more than one equation is given, they are used together to form a single design. For example, consider the resolution III, 2^{5-2} design. The equation D=+AB generates the points

A	B	D
-1	-1	+1
+1	-1	-1
-1	+1	-1
+1	+1	+1 .

Using the second equation E=+AC, one repeats the set for both positive and negative values of C and then determines the E values. The complete experimental design is

A	B	C	D	E
-1	-1	+1	+1	-1
+1	-1	+1	-1	+1
-1	+1	+1	-1	-1
+1	+1	+1	+1	+1
-1	-1	-1	+1	+1
+1	-1	-1	-1	-1
-1	+1	-1	-1	+1
+1	+1	-1	+1	-1 .

This concludes the brief introduction to the full factorial, the 2^{k-p} fractional factorial, and the central composite designs. These designs are used for two reasons. First, they require the minimum number of experiments to determine the information we need for our analysis of the response surface. Second, they are or can be formed into a larger class of

experimental designs known as orthogonal designs. An orthogonal design is a design with an X matrix that, when multiplied by its transpose, forms a diagonal matrix, $X'X = Z_{diag}$. (X matrices will be described in greater detail later.) These designs simplify the mathematics of RSM and improve the models generated.

THE METHOD OF LEAST SQUARES

The purpose of the method of least squares is to fit a polynomial to a response. For example, the three-parameter experiment that we have been using for our example could have the first-order polynomial equation

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon,$$

where y is the response; β_0, β_1, β_2, and β_3 are constant coefficients; x_1, x_2, and x_3 are coded values of a, b, and c, respectively; and ε is a random error accounting for our inability to find the true model. For this linear model we can write a matrix form that would include all the information of our experiments

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad [3]$$

where the vector $\mathbf{y}' = [\, y_1, y_2, \ldots, y_8]$ (the measured responses of our experiment), X is the previously defined X matrix (in this case, it is the linear model described in the previous section), $\boldsymbol{\beta}$ is a vector of the coefficients, $\boldsymbol{\beta}' = [\beta_0, \beta_1, \beta_2, \beta_3]$, and $\boldsymbol{\varepsilon}$ is a vector of errors. In this model, we know y and X but do not know $\boldsymbol{\beta}$ or $\boldsymbol{\varepsilon}$. The desired end is to determine the $\boldsymbol{\beta}$ vector for

3 The boldface letters used in these and the following equations denote vectors. The capital letters denote matrices. The size of these vectors and matrices is dependent on the size of the experiment and the complexity of the model. The ' denotes the transpose of a matrix or vector.

which the summation of the squares of the errors in the vector ε is a minimum, that is where

$$L = \sum_{i=1}^{8} \varepsilon_i^2 = \varepsilon'\varepsilon$$

is minimized. This can be written in the form

$$L = (\mathbf{y} - X\hat{\beta})'(\mathbf{y} - X\hat{\beta}) \quad [4]$$

Expanding the right side of the equation, we have

$$L = \mathbf{y}'\mathbf{y} - (X\hat{\beta})'\mathbf{y} - \mathbf{y}'X\hat{\beta} + (X\hat{\beta})'X\hat{\beta}$$
$$L = \mathbf{y}'\mathbf{y} - \hat{\beta}'X'\mathbf{y} - \mathbf{y}'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$
$$L = \mathbf{y}'\mathbf{y} - 2\hat{\beta}'X'\mathbf{y} + \hat{\beta}'X'X\hat{\beta}$$

The point at which L is a minimum as $\hat{\beta}$ is changed can be determined by differentiation. The minimum occurs where $\delta L/\delta\hat{\beta} = 0$. For the equation above,

$$dL/\delta\hat{\beta} = -2X'\mathbf{y} + 2(X'X)\hat{\beta}$$

Setting $\delta L/\delta\hat{\beta}$ to 0 and reducing, we have

$$(X'X)\hat{\beta} = X'\mathbf{y}$$
$$\hat{\beta} = (X'X)^{-1}X'\mathbf{y}.$$

4 The ^ in these equations denotes the term the regression predicts. The values of **y** and X were measured in the experiment. β we do not know and are attempting to predict by minimizing the scalar L. The ' in these equations denotes the transpose of a vector or matrix, so **a'b** is equivalent to the magnitude of a dot product, $\mathbf{a'b}=a_1b_1+a_2b_2+...+a_kb_k$.

Although this derivation of a least squares estimator was applied to a linear example, it is equally valid for higher order models. The type of model the least squares regression generates is solely dependent on the X matrix used in the regression. Therefore, it is important to understand how X matrices are formed.

The method of generating an X matrix is rather straightforward and simply requires the evaluation of the experimental parameters for each proposed coefficient in the model. The rows of the X matrix correspond to each of the points where experiments are conducted. The columns of the X matrix correspond to each of the terms the experimenter decides to include in his model. Neither the order of the experiments nor the order of the model terms affects the final expression. To determine the entry at the j^{th} column of the i^{th} row, the experimenter would evaluate the parameters of the i^{th} experiment in the relation defined by the j^{th} column. As an example, consider the entry of a complete row of a three-parameter experiment that is to be fitted to a model that includes a constant term, all main effects, and all two factor interactions. The model is to look like

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3$$

Consider the parameter of the i^{th} experiment to be $x_1 = 2$, $x_2 = 0$, and $x_3 = -1$. Then the first term of the row corresponds to β_0. Since β_0 is a constant term equivalent to the response at the center of the experimental design, it is always weighted as 1. This is the first entry. The remaining entries are calculated as follows:

$$x_1 = 2,$$

$$x_2 = 0,$$

$$x_3 = -1,$$

$$x_1x_2 = (2)(0) = 0,$$

$$x_1x_3 = (2)(-1) = -2,$$

$$x_2x_3 = (0)(-1) = 0.$$

So the i^{th} row would be

1 | 2 0 -1 0 -2 0.

To build the X matrix, this series of exercises is repeated for all the experiments.

X MATRICES FOR LINEAR, INTERACTION, AND SECOND ORDER MODELS

The three most frequently used models in process optimization are the linear, the interaction, and the second-order models. These models are used together with the experimental designs presented earlier. Below are examples of the X matrices for the linear, the interaction, and the second order models formed from these designs.

The generalized linear model of a three-parameter process is

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3$$

For the 2^{3-1} fractional factorial experiment described earlier, the X matrix would be

$$X = \left[\begin{array}{c|ccc} & x_1 & x_2 & x_3 \\ \hline 1 & +1 & -1 & -1 \\ 1 & -1 & +1 & -1 \\ 1 & -1 & -1 & +1 \\ 1 & +1 & +1 & +1 \end{array}\right].$$

The model generated by the least squares regression is written in the form

$$\mathbf{y} = b_0 + \mathbf{x'b}$$

with

$$b_0 = \beta_0,$$

$$\mathbf{x}' = [x_1, x_2, x_3],$$

and

$$\mathbf{b} = \begin{bmatrix} \beta 1 \\ \beta 2 \\ \beta 3 \end{bmatrix}.$$

The interaction model considers the average response, all main effects, and all possible interactions. The model for a three-parameter process is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_3 x_3 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3.$$

The smallest experimental design that will generate this model is a full factorial design. If a fractional factorial design is used, confounding occurs. Below is the corresponding X matrix for the interaction model of a 2^3 full factorial experiment.

$$X = \left[\begin{array}{c|ccccccc} & x1 & x2 & x1x2 & x3 & x1x3 & x2x3 & x1x2x3 \\ \hline 1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 \\ 1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ 1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ 1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ 1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\ 1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\ 1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{array}\right].$$

The second-order model considers the average response, the main effects, the squared effects, and the two-factor interactions. The model for a three-parameter process is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11}(x_1)^2 + \beta_{22}(x_2)^2 + \beta_{33}(x_3)^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 .$$

The smallest experimental design, of the ones described, that will generate this model is a resolution **V** or better fractional factorial design combined with the center and axial points of a central composite design. Since no three-parameter resolution **V** fractional factorial designs exist, a full central composite design must be used in this case. The X matrix for the second-order model of a three-parameter process using a central composite design is

$$X = \begin{array}{c} \\ \left[\begin{array}{c|ccccccccc} & x_1 & x_2 & x_3 & (x_1)^2 & (x_2)^2 & (x_3)^2 & x_1x_2 & x_1x_3 & x_2x_3 \\ \hline 1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 & +1 & +1 \\ 1 & +1 & -1 & -1 & +1 & +1 & +1 & -1 & -1 & +1 \\ 1 & -1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & -1 \\ 1 & +1 & +1 & -1 & +1 & +1 & +1 & +1 & -1 & -1 \\ 1 & -1 & -1 & +1 & +1 & +1 & +1 & +1 & -1 & -1 \\ 1 & +1 & -1 & +1 & +1 & +1 & +1 & -1 & +1 & -1 \\ 1 & -1 & +1 & +1 & +1 & +1 & +1 & -1 & -1 & +1 \\ 1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ 1 & -1.216 & 0 & 0 & 1.479 & 0 & 0 & 0 & 0 & 0 \\ 1 & +1.216 & 0 & 0 & 1.479 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1.216 & 0 & 0 & 1.479 & 0 & 0 & 0 & 0 \\ 1 & 0 & +1.216 & 0 & 0 & 1.479 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1.216 & 0 & 0 & 1.479 & 0 & 0 & 0 \\ 1 & 0 & 0 & +1.216 & 0 & 0 & 1.479 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array} .$$

At this point, a digression is in order. Note the selection of the coded value of 1.216 for the axial distances. The value was selected to retain an orthogonal design. Orthogonal designs are important in a least squares

regression since they are rotatable. A rotatable design is one in which the least squares model predicts a response that has an estimated variance that is only dependent on the distance from the design center. Most research conducted on experimental methodology is directed at developing designs that retain this property. The value 1.216 comes from a method presented by Meyers [5:127-134] .

$$\alpha = \left(\frac{QF}{4} \right)^{1/4}$$

where $F=2^{k-p}$, the number of factorial points, and Q is a function of the F and another value T. T = 2k + z, the number of axial and center points, by the relation

$$Q = [(F + T)^{1/2} - F^{1/2}]^2 .$$

Box and Draper present another more general method for choosing α for these designs summarized in the equation

$$\alpha = \left(\frac{2^{k-p} r_c}{r_s} \right)^{1/4}$$

where 2^{k-p} is the number of experimental points, r_c is the number of replicates of the factorial experiments (repetitions of the 2^{k-p} design) and r_s is the number of replicates of the axial experiments [1:477-524]. The α's predicted by these two methods are different. The α of 1.216 is determined using Meyer's method. Since it is not the intent of this thesis to compare the merits of these two methods, the interested reader should refer to the sources [1] and [5].

The equation of the second-order model may be written in the form

$$\hat{y} = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}.$$

The terms in the equation are defined as

$$b_0 = \beta_0,$$

$$\mathbf{x}' = [x_1, x_2, x_3],$$

$$\mathbf{b} = \begin{bmatrix} \beta 1 \\ \beta 2 \\ \beta 3 \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} \beta 11 & \beta 12/2 & \beta 13/2 \\ \beta 12/2 & \beta 22 & \beta 23/2 \\ \beta 13/2 & \beta 23/2 & \beta 33 \end{bmatrix}.$$

As you will see in Chapter 4, the second-order model is extremely important in response surface studies .

A STATISTICAL ANALYSIS OF THE REGRESSION MODEL

Once an experimenter obtains a regression model for a response, he should ask the question: "What confidence can I have that the model represents the process?" In a least squares regression, the sum of the squares of the errors, SS_E, is minimized; however, if the minimized SS_E remains large, it will reduce the experimenter's confidence in the model's adequacy. An inquisitive experimenter would want to know further whether the SS_E is due to the pure error or to the model's lack of fit to the true process. This section gives a brief explanation of the procedures used to answer these questions.

In regression analysis it is assumed that the expected values of the errors are zero and that they are distributed about this expected value with a constant variance of σ. If such an assumption is correct, then it can be shown

that the expected value of the vector β is $\hat{\beta}$.[6] In such a situation, $\hat{\beta}$ is considered an unbiased estimator of the response, and the covariance of its estimates are a function of the constant variance of the error. The following equations show these assumptions.

ASSUMPTION: $E(\varepsilon) = 0,\ Cov(\varepsilon) = \sigma^2$

THEN:

$$\begin{aligned} E(\hat{\beta}) &= E[(X'X)^{-1}X'\mathbf{y}] \\ &= E[(X'X)^{-1}X'(X\beta + \varepsilon)] \\ &= E[I\beta + (X'X)^{-1}X'\varepsilon] \\ &= E[\beta] + E[\varepsilon] \\ &= \beta \end{aligned}$$

$$\begin{aligned} Cov(\mathbf{y}) &= Cov[X\beta + \varepsilon] \\ &= Cov[X\beta] + Cov[\varepsilon] \\ &= 0 + I\sigma^2 \end{aligned}$$

6 The convention used here is that the ^ defines terms of the model generated by the regression. $\mathbf{y}$ is the vector of measured responses and $\hat{\beta}$ is the vector in the assumed model $\mathbf{y} = X\beta + \varepsilon$. $\hat{\mathbf{y}}$ is the vector of responses predicted by the least squares model $\hat{\mathbf{y}} = X\hat{\beta}$.

$$\begin{aligned} \mathrm{Cov}(\hat{\beta}) &= \mathrm{Cov}[(X'X)^{-1}X'\mathbf{y}] \\ &= ((X'X)^{-1}X')'((X'X)^{-1}X')\ \mathrm{Cov}[\mathbf{y}] \\ &= (X'X)^{-1}\ \mathrm{Cov}[\mathbf{y}] \\ &= (X'X)^{-1}\ \sigma^2 \end{aligned}$$

$$\begin{aligned} \mathrm{Cov}(\hat{\mathbf{y}}) &= \mathrm{Cov}[\mathbf{x}'\hat{\beta}] \\ &= \mathbf{x}'\mathbf{x}\ \mathrm{Cov}[\hat{\beta}] \\ &= \mathbf{x}'\mathbf{x}\,(X'X)^{-1}\ \sigma^2 \end{aligned}$$

The last function illustrates how an orthogonal design is also a rotatable design. When X'X is a diagonal matrix with equal diagonal entries, q, the variance of $\hat{y}$, is simply a function of the distance from the design center,

$$\begin{aligned} \mathrm{Cov}[\hat{y}] &= \mathbf{x}'\mathbf{x}\ \frac{1}{q} I\ \sigma^2 \\ &= \mathbf{x}'\mathbf{x}\ \frac{1}{q}\ \sigma^2 \ . \end{aligned}$$

The fact that $\hat{\beta}$ is an estimate of β implies that the variances of the responses predicted by the least squares model are the same as the variances of the responses of the process. This is the foundation of the analysis of variance. If the variance of the responses predicted by the model are significantly different than the variance of the experiments, then we cannot have confidence that the model accurately represents the process.

As a foundation of how the variance is analyzed, it is helpful to review basic concepts of statistics. Let us begin by defining three values, sum of squares total (SS_T), sum of squares error (SS_E), and sum of squares average (SS_A). All are related by the equation

$$SS_T = SS_A + SS_E$$

Now, consider a series of n experiments conducted using a single set of parameter settings with responses y_i, (i = 1,2,...,n). The SS_T is defined as the sum of squares of all the measured responses

$$SS_T = \sum_{i=1}^{n} y_i^2 \quad .$$

The SS_A is defined as n times the square of the averages of all the reponses

$$SS_A = n \left(\frac{\sum_{i=1}^{n} y_i}{n} \right)^2 = n \bar{y}^2$$

Then, by using the equality defined first, we know that

$$SS_E = SS_T - SS_A$$

so

$$SS_E = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2,$$

which reduces to

$$SS_E = \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad . \quad (*)$$

This is consistent with what one expects for the SS_E. The average value of a response of a series of experiments is always considered more representative of the response's value than any of the values of the

responses measured in the experiments. Therefore, the difference of a measured response from an average response is considered an error,

$$\text{Error}_i = y_i - \bar{y} \quad .$$

The sum of squares of these errors is

$$SS_E = \sum_{i=1}^{n} \text{Error}_i^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 ,$$

which is the same expression as equation (*).

Now, there is a second series of values known as the mean squares. Mean squares, as the name suggests, are a mean of the sum of squares. Clearly, the mean of the SS_T is SS_T/n, but, the mean squares of the two terms SS_E and SS_A are not as obvious. Another way of looking at the mean squares is as a normalized sum of squares that accounts for the number and type of experiments evaluated. To account for the number and type of experiments, we use a third term called the degrees of freedom (DOF). By convention, the value n is referred to the degrees of freedom in the experiment. The degrees of freedom for each SS term must follow the same equality as the SS equality above. So,

$$DOF_T = DOF_E + DOF_A,$$

and, in this case,

$$DOF_T = DOF_E + DOF_A = n.$$

The SS_A term can be written as

$$SS_A = (1/n)\left(\sum_{i=1}^{n} y_i \right)^2 .$$

In a way, the SS_A is already an average of a single squared value and, therefore, only one degree of freedom is associated with it, so the mean square average, MS_A, is

$$MS_A = \frac{SS_A}{1} .$$

This leaves n-1 degrees of freedom for the calculation of the mean square error, MS_E,

$$MS_E = \frac{SS_E}{n-1} .$$

Anyone familiar with statistics will immediately recognize that this is the sample variance which is often used to estimate the population variance.

The concepts presented here extend to the analysis of the variance of the least squares model for which the sum of squares equality is

$$SS_T = SS_{ERROR} + SS_{regression} .$$

The latter term can be further broken down into two components,

$$SS_{regression} = SS_{\beta 0} + SS_{(R|\beta 0)} .$$

$SS_{\beta 0}$ is the portion of the $SS_{regression}$ related to the term β_0 and $SS_{(R|\beta 0)}$ is the portion related to the rest of the model. Recalling that β_0 was the biasing term equal to the response at the center of the experiment, it follows that $SS_{\beta 0}$ is similar to the SS_A. So by comparison to our introducton,

$$SS_E = SS_{ERROR} + SS_{(R|\beta 0)} .$$

When we decided to perform a least squares regression, we already knew that the SS_E was large. Qualitatively, what we were trying to do in a least squares regression was to put as much as possible of the SS_E term into

the $SS_{(R|\beta 0)}$ and to minimize the SS_{ERROR} term. The measure of our success is seen in the ratio of these values. If the $SS_{(R|\beta 0)}$ is much larger than the SS_{ERROR} then our regression is successful.

Now we need to calculate these sum of squares and mean squares. As before, the SS_T is the sum of the squares of all the responses

$$SS_T = \sum_{i=1}^{n} y_i^2 = \mathbf{y'y}.$$

The $SS_{regression}$ is determined similarly, but evaluates the squares of the predicted responses

$$\begin{aligned} SS_{regression} &= \sum_{i=1}^{n} \hat{y}_i^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}} \\ &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \ . \end{aligned}$$

As stated above, the $SS_{regression}$ term has two parts, $SS_{\beta 0}$ and $SS_{(R|\beta 0)}$ with $SS_{\beta 0}$ being equivalent to the SS_A. By induction, it follows that the $SS_{\beta 0}$ is determined from the average of the responses of the experiment (the same as the SS_A),

$$SS_{\beta 0} = n \left(\frac{\sum_{i=1}^{n} y_i}{n} \right)^2 = n\,\overline{y}^2 = (1/n)\mathbf{y'y} \ .$$

In turn, it follows that the difference between the overall regression sum of squares and the $SS_{\beta 0}$ term provides the $SS_{(R|\beta 0)}$ term,

$$SS(R|\beta_0) = \hat{\beta}'X'y - n\bar{y}^2$$

Finally, we have to determine the error sum of squares. The obvious method is to evaluate the difference between SS_T and $SS_{regression}$,

$$SS_{ERROR} = SS_T - SS_{regression} = \mathbf{y'y} - \hat{\beta}'X'\mathbf{y}.$$

But, there is another way to approach this that will enable us to obtain a better understanding of the error sum of squares, SS_{ERROR}. The SS_{ERROR} is often referred to as the sum of squares of the residuals, the $SS_{residual}$. A residual is the difference between the response predicted by the model and the actual experimentally measured response,

$$residual_i = y_i - \hat{y}_i \quad (i = 1,2,...,n) .$$

So, the SS_{ERROR} is defined as

$$SS_{ERROR} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 .$$

This is equivalent to

$$\begin{aligned} SS_{ERROR} &= (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) \\ &= \mathbf{y'y} - \beta'X'\mathbf{y} \quad , \end{aligned}$$

which is the same expression for the SS_{ERROR} derived in the first method.

The determination of the degrees of freedom (DOF) is also similar. Again, n degrees of freedom are associated with the SS_T. The degrees of freedom associated with regression are equivalent to the number of

coefficients in the model, k. These k degrees of freedom are split between the two regression terms, with 1 going to the term, $SS_{\beta 0}$, and k-1 going to the $SS_{(R|\beta 0)}$. Conserving our degrees of freedom, we know that

$$DOF_T = DOF_{ERROR} + DOF_{regression}$$

so

$$DOF_{ERROR} = DOF_T - DOF_{regression} = n-k.$$

Once we obtain both the sum of squares terms and the degrees of freedom associated with them, we can calculate the mean square terms using the general relation MS = SS/DOF. Table 2.2 is a summary of all the calculations, known as the Analysis of Variance Table, leading up to our ability to calculate these mean squares.

Table 2.2 **The ANALYSIS OF VARIANCE**

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE
REGRESSION β_0	$n\bar{y}^2$	1	$SS(\beta_0)$
$R\|\beta_0$	$\hat{\beta}'X'y - n\bar{y}^2$	k - 1	$SS(R\|\beta_0)/k$
ERROR	$y'y - \hat{\beta}'X'y$	n - k	$SS_E/(n - k)$
TOTAL	$y'y$	n	

k is the number of parameters in the model; n is the number of experiments.

The measure of the adequacy of a model can be seen in the ratio of the variance of the regression terms to the variance of the error term. If the variance of the regression term is significantly different than (larger than) the

variance of the error, then we can state that the regression was good; the variation of the response as a function of the parameter settings is predicted well by the model. The statistical test for the significance of our regression is the hypothesis test using the F statistic.[7] In this test, we start with the hypothesis that the mean squares of the regression and error are variances of distributions with the same expected value,0. The fact that the error has an expected value of 0 is a fundamental assumption of a least squares regression. Therefore, the MS_{ERROR} term is considered the true variance of the distribution that has 0 as the expected value. In the F test, one evaluates the expected values of the regression terms by comparing their variance, their mean square term, to the variance of the error, MS_{ERROR} For example, let us evaluate the constant regression term β_0. To determine if the constant term is significant, we test the hypothesis

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta_0 \neq 0$$

For φ confidence that the β_0 term is significant (we can reject the hypothesis $\beta_0 = 0$), then the following must be true

$$F = MS(\beta_0)/ MS_E \geq F_{1,n-k,1-\varphi} .$$

In most cases, we expect the β_0 term to be significant. The true test for the regression's adequacy is seen in whether the remaining regression terms are

7 The F statistic is used to check a hypothesis concerning the ratio of population variances of two normal distributions. Virtually all texts on statistics have a thorough presentation of how they are used and have tables which can be referenced for the analysis.

significant, since they are the portion of the model that define the variation of the response from the average. So, we apply the same technique to test the significance of the remaining regression terms. To determine if the remaining regression terms are significant, we test the hypothesis

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

$$H_1 : \beta_i \neq 0 \text{ for some } i \quad .$$

For φ confidence that they are significant, then the following must be true

$$F = MS(R|\beta_0)/MS_E \geq F_{k,n-k,1-\varphi}.$$

There is one fallacy in these tests. The error variance does not represent the error variance of the process, but is a combination of the random error of the process and the error in the model due to its lack of fit. If the lack of fit is dominant, the model may be inadequate for our use, even though the F statistic of the regression indicated its adequacy. Conversely, if the random error is dominant, we can never conclude that the type of model is inadequate but only that the specific model is inadequate.

To test for the lack of fit, one must separate the error into two components. An estimate of the pure error can be made by replicating experiments at one or more of the data points.[8] If r_i replicates are made at the i^{th} data point, its contribution to the $SS_{PUREERROR}$ is

$$\sum_{u=1}^{r_i} (y_{iu} - \bar{y}_i)^2 \quad .$$

8 Recall, a data point is a location in the experimental design that is defined by a unique set of parameters.

This equation is the same as equation (*), the equation for evaluating the SS_E when only one data point is considered. The calculation of the $SS_{PURE\,ERROR}$ is simply the summation of these contributions over all m data points, where

$$SS_{PURE\ ERROR} = \sum_{i=1}^{m} \sum_{u=1}^{r_i} (y_{iu} - \bar{y}_i)^2 \quad .$$

The degrees of freedom for the estimate can be determined in a similar manner. The contribution of the experiments at the i^{th} data point to the $DOF_{PURE\,ERROR}$ is

$$(r_i - 1).$$

In turn, the $DOF_{PURE\,ERROR}$ is simply the summation of these contributions over all m data points, where

$$DOF_{PURE\ ERROR} = \sum_{i=1}^{m} (r_i - 1) = \sum_{i=1}^{m} r_i - m \quad .$$

The sum of squares for the lack of fit can be defined as the difference of the pure error sum of squares from the residual sum of squares,[9]

$$SS_{LOF} = SS_{residual} - SS_{PURE\ ERROR} \quad .$$

9 Recall that the residual sum of squares is the same as the SS_{ERROR} All the SS relations are summarized below to assist the reader in avoiding confusion.

$$SS_T = SS_E + SS_A,$$
$$SS_T = SS_{ERROR} + SS_{regression},$$
$$SS_{regression} = SS_{\beta 0} + SS(R|\beta_0),$$
$$SS_{\beta 0} = SS_A$$
$$SS_E = SS_{ERROR} + SS(R|\beta_0),$$
$$SS_{ERROR} = SS_{residual},$$
$$SS_{residual} = SS_{PURE\ ERROR} + SS_{LOF}.$$

To develop an expression for the SS_{LOF} it is necessary to evaluate the residual of the u^{th} observation at the i^{th} data point,

$$y_{iu} - \hat{y}_i = (y_{iu} - \bar{y}_i) - (\hat{y}_i - \bar{y}_i) \quad ,$$

the difference between the measured value of the u^{th} observation at the i^{th} data point, and the predicted value for the i^{th} data point. Squaring both sides of this equation and summing over all i's and u's can provide two equivalent expressions for the residual sum of squares:

$$\sum_{i=1}^{m} \sum_{u=1}^{r_i} (y_{iu} - \hat{y}_i)^2 = \sum_{i=1}^{m} \sum_{u=1}^{r_i} (y_{iu} - \bar{y}_i)^2 + \sum_{i=1}^{m} r_i (\hat{y}_i - \bar{y}_i)^2 .$$

Notice that the first term on the right half of the expression is equivalent to the pure error sum of squares. Since

$$SS_{residual} = SS_{PURE\ ERROR} + SS_{Lack\ of\ Fit} \quad ,$$

it follows that the expression for the lack of fit sum of squares is

$$SS_{LOF} = \sum_{i=1}^{m} r_i (\hat{y}_i - \bar{y}_i)^2 \quad .$$

The degrees of freedom for the $SS_{residual}$ can be evaluated by the expression

$$DOF_T = DOF_{residual} + DOF_{regression} \quad .$$

Since DOF_T is the total number of experiments, n, and the $DOF_{regression}$ is the number of coefficients in the model, k, the $DOF_{residual}$ is simply the total number of experiments minus the degrees of freedom associated with the model coefficients,

$$DOF_{residual} = n - k \quad .$$

Since we now know the $DOF_{residual}$ and the $DOF_{PURE\ ERROR}$, we can determine the DOF_{LOF} using the relation

$$DOF_{LOF} = DOF_{residual} - DOF_{PURE\ ERROR} \quad .$$

Rewriting n as the summation of all r_i, the DOF_{LOF} can be determined readily.

$$DOF_{LOF} = \sum_{i=1}^{m} r_i - k - \left(\sum_{i=1}^{m} r_i - m \right) = m - k \quad .$$

Table 2.3 is the revised analysis of variance table.

Table 2.3 ANALYSIS OF VARIANCE when experiments are replicated

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE
REGRESSION β_0	$n\bar{y}^2$	1	$SS(\beta_0)$
R| β_0	$\hat{\beta}'X'y - n\bar{y}^2$	k - 1	$SS(R\|\beta_0)/(k-1)$
LACK OF FIT	$\sum_{i=1}^{m} r_i(\hat{y}_i - \bar{y}_i)^2$	m - k	$SS_{LOF}/(m - k)$
PURE ERROR	$\sum_{i=1}^{m} \sum_{u=1}^{r_i} (y_{iu} - \bar{y}_i)^2$	n - m	$SS_E/(n - m)$
TOTAL	$\sum_{i=1}^{m} \sum_{u=1}^{r_i} y_{iu}^2$	n	

At this point it may be helpful to explain what the values n, m, k, and r_i stand for. n is the total number of experiments performed; m is the total number of data points evaluated in the designed experiment; k is the number of coefficients in the least squares model; and r_i is tne number of replicates of the experiment at datapoint i. The relationship of n, m, and r_i to each other is summarized in the equation

$$n = \sum_{i=1}^{m} r_i \quad .$$

To determine if the lack of fit is significant, we test the hypothesis

$$H_0 : \text{Lack of fit} = 0$$

$$H_1 : \text{Lack of fit} \neq 0.$$

For φ confidence that the lack of fit is significant, the following must be true

$$F_{LOF} = MS_{LOF}/MS_E \geq F_{m-k,n-m,1-\rho}$$

For large values of F_{LOF}, the adequacy of our model is discredited.

When the experimenter encounters a large F_{LOF}, he must decide if the model is adequate for his studies. Before redoing any experiments or using a different design, the experimenter should attempt to achieve a better fit by transforming the response, by transforming one or more of the variables, or by using a completely different model. (All models are wrong; the issue is how wrong.) Even with a large F_{LOF}, our model may still be useful in our application of the RSM.

THE MODEL ADEQUACY AND THE RECODING OF VARIABLES

For large values of F_{LOF}, the adequacy of the model is discredited but this is not necessarily an indication that the data is incorrect. Oftentimes the coding of the parameters, the units of the response, the coding of the response, or a combination of these may be wrong. A subsequent recoding of the variables or the response may improve the fit. As an example of the recoding one parameter, consider the coding of the "a" parameter in our past example:

SETTINGS (a)	CODED VALUES (x_1)
3.9	-1.216
5.0	-1.000
10.0	0.000
15.0	+ 1.000
16.1	+ 1.216 .

A possible recoding could use the natural logarithm of the parameter. To retain the coded values -1 and +1 for the settings 5 and 15, we determine a new $\bar{p}$ and d using ln(5) and ln(15). (The -1 and +1 codings are normally kept the same to keep the design somewhat in a rotatable form.)

$$\bar{p} = (\ln(5) + \ln(15))/2 = 2.159$$

$$d = (\ln(15) - \ln(5))/2 = .549$$

Then one can easily determine the coded parameters for 3.9, 10, and 16.1.

The recoded parameters are

for 3.9, (ln(3.9)-2.159)/.549 = -1.454,

for 5.0, (ln(5.0)-2.159)/.549 = -1.000,

for 10.0, (ln(10)-2.159)/.549 = 0.262,

for 15.0, (ln(15)-2.159)/.549 = 1.000,

and for 16.1, (ln(16.1)-2.159)/.549 = 1.129.

After recoding the parameters, the entries in the X matrix used to form the model are changed, and the least squares regression is repeated for a new model.

These methods of recoding are "best guess" operations which the experimenter would execute based on his knowledge of the process or his intuition. I am not aware of any algorithms that can be used to optimize coding. Recoding may or may not improve the model (generate a lower F statistic for the lack of fit), and the experimenter will be faced with the dilemma of either scrapping the experiment, using a different model, or using a poor model for further analysis. Since conducting experiments is usually expensive, the first alternative is rarely a good one. If the experimenter is trying to build a model to simply better understand the process, the second alternative of using a different model would probably be best. But, if the purpose is to optimize a process, even a poor second-order model is preferable to a higher order model. The reason for this is that the analysis of a second-order model provides us with the most direct tools for a response surface analysis.

CHAPTER 3

THE ANALYSIS OF EXPERIMENTS FOR PARAMETER SIGNIFICANCE

The study of any process starts with an analysis of the parameters that are controlled or varied in the process. The experimenter tries to determine which parameters are significant to the response(s) of the system. In the first step of the analysis, the researcher conducts a preliminary investigation of the process, identifies the possible parameters, and eliminates those which he is fairly certain are insignificant. If the researcher is not certain whether any of the remaining processes are significant, then his next step is to perform an experiment. In the early days of research this next step would have involved the analysis of experiments that simply varied the questionable variable. In many cases, such a procedure is adequate, but, an inherent error exists in the procedure. It does not account for the interactions of other parameters which may hide the questioned parameter's main effect. Additionally, it does not evaluate how the questioned parameter interacts with the other parameters. Experimental methodology has evolved since those early days, and today we use a group of experimental designs and analysis methods that allow us to evaluate these parameter effects. The designs are the factorial and fractional factorial designs explained in Chapter 2. And, the analysis simply involves an ordered application of the interaction model, also presented in Chapter 2.

HISTORICAL DEVELOPMENT

The field of statistics grew out of the agricultural research of R. A. Fisher in the early part of this century. His research dealt with the effects of

discrete changes in the environment on plant growth. As a result, experimental designs were developed to answer questions of a discrete nature as opposed to a continuous nature. As an illustration, consider the difference between determining the best of two brands of fertilizer and determining the best concentration of nitrogen in a fertilizer. The first illustration involves two discrete possibilities while the latter involves a continuous range of possibilities. It is interesting to think that the full factorial experiment developed from this perspective as opposed to the continuous perspective that makes it so well suited for a least squares regression. Unfortunately, the discrete perspective still remains the perspective that most texts use to introduce the subject. There is a good reason for this choice. Discrete criteria are still factors in most processes. The disadvantage of the perspective is the difficulty in explaining it. Consider a simple two-factor experiment with factors a and b.[1] In a full factorial design, the effect A is simply evaluated as the average of the differences of the responses of the experiment when B is constant and A changes

$$\frac{((A^+ B^-) - (A^- B^-)) - ((A^+ B^+) - (A^- B^+))}{2}$$

1 In Chapter 3, the uppercase letter is used interchangeably to refer to the high and low parameter settings, to refer to responses, and to refer to effects. This is the standard convention found in all texts. If one considers the context in which the letters are used, there should be no confusion. When used in an equation, the letter refers to the high and low parameter settings. The statement $A^+ B^-$ means the response when A is high and B is low. The interaction AB is the effect of the interaction between the A parameter and the B parameter.

The evaluation of the effect B is similar,

$$\frac{((A^+ B^+) - (A^+ B^-)) - ((A^- B^+) - (A^- B^-))}{2}$$

But, the evaluation of the interaction of the A and B parameters, AB, is less obvious. It is the difference of the averages when A and B are both opposites,

$$\frac{((A^+ B^+) + (A^- B^-))}{2} - \frac{((A^+ B^-) - (A^- B^+))}{2}$$

The equation quickly becomes more complicated as the size of an experiment increases and as larger interactions are considered. Most authors of textbooks in the subject recognize the difficulty and normally include a table of +'s and – 's to show the evaluation of the responses for a particular interaction. Most researchers, in turn, resort to using these tables to evaluate their experiments.

In the late 1930's, Frank Yates introduced an algorithm that greatly simplified the calculations of the effects. His algorithm simply involved organizing the responses in a column and repeating an algorithm of adding and subtracting adjacent entries. Although this empirically derived procedure is effective at simplifying the mechanics of the calculations, it does little to explain the relations of the responses or to explain why some are added and others are subtracted. For this explanation, it is necessary to return to the subject of least squares regression. Upon examination of the analysis of full factorial experiments, one finds that the relations of the responses are the same as those generated when least squares regression is used to form an interaction model. The sign of the contributions of each response to the estimate of an effect is nothing more than that determined when we filled in

the column entries of the X matrix of the interaction model in Chapter 2. So the values of the effects and interactions are related to the coefficients of the interaction model (twice the coefficient). Yates' algorithm simply produces these values when the experiments (rows of the X matrix) are organized in a particular order.

It is helpful to recognize that the purpose of Yates' algorithm is not to build a model of a process but to evaluate the significance of the effects and the interactions of the parameters of the process. The purpose of least squares and Yates' algorithm are different and, therefore, are rarely compared. However, understanding the method of least squares makes the comprehension of Yates' algorithm much easier. Therefore, I will use the unconventional approach of covering Yates' algorithm by comparing it to the least squares interaction model.

YATES' ORDER

The ordering of experiments in Yates' algorithm can best be compared to counting in binary numbers. The binary number is formed with the high and low states, +1 and -1, of all the parameters considered in the experiment. By convention, the least significant digit of this number is in the leftmost column.[2] So , the order of the experiments in 2^2 and 2^3 factorial experiments is

2 A standard binary number is formed with the two digits, 1 and 0. As an example, consider the binary equivalent of the decimal number 13, 1101. In Yates' ordering, the two binary digits are +1 and -1. Rewriting 13 as a binary number with these two digits provides +1 +1 -1 +1. If the left most digit becomes the least significant digit, 13 is written as +1-1+1+1.

	A	B
1	−1	−1
2	+1	−1
3	−1	+1
4	+1	+1

and

	A	B	C
1	−1	−1	−1
2	+1	−1	−1
3	−1	+1	−1
4	+1	+1	−1
5	−1	−1	+1
6	+1	−1	+1
7	−1	+1	+1
8	+1	+1	+1

Yates' algorithm is a repetitive procedure which generates a column of values that are the effects of the different parameters and their interactions. The order of the effects that Yates' algorithm generates can be determined directly from the order of the experiments. The effect that corresponds to each experiment is equivalent to the positive parameter values. For example, the effect that corresponds to

A	B	C
−1	+1	−1

is B and the effect that corresponds to

A	B	C
−1	+1	+1

is BC. So the order of all the coefficients in our two examples is

	A	B	Effect	coef.
1	−1	−1	1	β_0
2	+1	−1	A	β_1
3	−1	+1	B	β_2
4	+1	+1	AB	β_{12}

	A	B	C	Effect	coef.
1	−1	−1	−1	1	β_0
2	+1	−1	−1	A	β_1
3	−1	+1	−1	B	β_2
4	+1	+1	−1	AB	β_{12}
5	−1	−1	+1	C	β_3
6	+1	−1	+1	AC	β_{13}
7	−1	+1	+1	BC	β_{23}
8	+1	+1	+1	ABC	β_{123}

3 The effect 1 is the average of the responses, or the response at the center of the axis system.

Using this order for the experiments and the model coefficients, the X matrix for the 2^3 factorial design is

EFFECT	1	A	B	AB	C	AC	BC	ABC
COEFFICIENT	β_0	β_1	β_2	β_{12}	β_3	β_{13}	β_{23}	β_{123}
GENERATOR		x_1	x_2	x_1x_2	x_3	x_1x_3	x_2x_3	$x_1x_2x_3$

$$X = \begin{bmatrix} 1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 \\ 1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ 1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ 1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ 1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\ 1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\ 1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{bmatrix}.$$

A least squares regression with this X matrix would generates the $\hat{\beta}$ vector

$$\hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{12} \\ \beta_3 \\ \beta_{13} \\ \beta_{23} \\ \beta_{123} \end{bmatrix}.$$

The order of the coefficients in the vector is the same as the order of the values for the effects and coefficients generated by Yates' algorithm.

COMPARING LEAST SQUARES REGRESSION TO YATES' ALGORITHM

If we took this matrix and the responses of the 8 experiments and applied a least squares regression, we would eventually find that the final value for each coefficient would be 1/8 the value determined by adding and

subtracting the responses in the manner specified by the rows of the transpose of the X matrix.[4] These additions and subtractions are the same as those specified by the columns beneath each effect. For example the column beneath AB is

$$\begin{array}{r} +1 \\ -1 \\ -1 \\ +1 \\ +1 \\ -1 \\ -1 \\ +1. \end{array}$$

Applying this in reverse order to the y vector we would then determine β_{12} as

$$\beta_{12} = \frac{(y_1 - y_2 - y_3 + y_4 + y_5 - y_6 - y_7 + y_8)}{8}$$

Consider how Yates' algorithm arrives at the same results. The mechanism of Yates' algorithm is simple. The responses of the experiment are listed in a column in the order of the experiments described earlier. A new column is then formed by making two passes down the first column and operating on adjacent pairs. In the first pass, the adjacent pairs are added to form the first 2^{k-1} entries of the new column. In the second pass, adjacent pairs are subtracted, the first from the second, to form the second 2^{k-1} entries of the new column. The algorithm is then applied a second time, this time to the new column to form still another column. The algorithm is repeated a total of

4 The matrix resulting from $(X'X)^{-1}$ is a diagonal matrix with 1/8 as the value for all diagonal entries, or more simply 1/8 I, I being the identity matrix. The regression equation reduces to $\beta = 1/8\ X'y$. This is true for all full factorial experiments. The value of the denominator of the fraction is 2^k, or more generally, the number of design points evaluated.

k times. As an example, consider our three parameter experiment. Below is the initial column layout of the responses and the subsequent values after each pass of Yates' algorithm.

y_1	$+y_1+y_2$	$+y_1+y_2+y_3+y_4$	$+y_1+y_2+y_3+y_4+y_5+y_6+y_7+y_8$	1
y_2	$+y_3+y_4$	$+y_5+y_6+y_7+y_8$	$-y_1+y_2-y_3+y_4-y_5+y_6-y_7+y_8$	A
y_3	$+y_5+y_6$	$-y_1+y_2-y_3+y_4$	$-y_1-y_2+y_3+y_4-y_5-y_6+y_7+y_8$	B
y_4	$+y_7+y_8$	$-y_5+y_6-y_7+y_8$	$+y_1-y_2-y_3+y_4+y_5-y_6-y_7+y_8$	AB
y_5	$-y_1+y_2$	$-y_1-y_2+y_3+y_4$	$-y_1-y_2-y_3-y_4+y_5+y_6+y_7+y_8$	C
y_6	$-y_3+y_4$	$-y_5-y_6+y_7+y_8$	$+y_1-y_2+y_3-y_4-y_5+y_6-y_7+y_8$	AC
y_7	$-y_5+y_6$	$+y_1-y_2-y_3+y_4$	$+y_1+y_2-y_3-y_4-y_5-y_6+y_7+y_8$	BC
y_8	$-y_7+y_8$	$+y_5-y_6-y_7+y_8$	$-y_1+y_2+y_3-y_4+y_5-y_6-y_7+y_8$	ABC

The final values in the last row are referred to as the contrasts of each of the effects. Dividing the contrasts by the total number of experiments executed gives the coefficients of the interaction model. By comparison one finds that the values generated by Yates' algorithm for the coefficients are identical to the corresponding coefficients generated by the least squares regression.

In the interaction model, the coefficients designate the sign and magnitude that the response varies from an average value, β_0, as a function of the particular interactions each precedes. When we use Yates' algorithm, we are actually looking for something else: the effect. The effect is the sign and magnitude that the response changes when moving from the low to the high state of the particular interaction. In these experiments, it is simply twice the value of the coefficients.[5]

5 Consider the term $\beta_1 x_1$ of an interaction model. The effect A is the contribution of this term when x_1 changes from -1 to + 1, so

$$A = \beta_1(+1) - \beta_1(-1),$$
$$A = 2\beta_1.$$

STATISTICAL CONSIDERATIONS AND EFFECT SIGNIFICANCE

The problem with applying the interaction model to a single replicate full factorial experiment is that the number of coefficients in the model is equal to the number of experiments. This presents a problem; there are no degrees of freedom available to allow any analysis of variance. A partial solution is to repeat some of the experiments, allowing an analysis of the regression, but not an analysis of the lack of fit. The additional experiments only increase the degrees of freedom associated with the experiment and the pure error. Luckily, the goal in the application of Yates' algorithm is not to analyze any particular model, but to analyze the regression or, more specifically, to analyze the regression of each of the coefficients. Therefore, replicating the experiments is a valid solution. The goal is to determine which coefficients are significant by comparing the mean square of the effects to the mean square of the pure error.

Yates' algorithm provides an organized method to analyze the effects of the parameters of a process provided some experiments are repeated. The normal practice is to replicate all the experiments in the factorial design, sum the responses for each experimental point, and then apply Yates' algorithm. This would provide a whole new set of contrasts that would be larger than the previous ones by a factor of n, the number of replicates used. From these contrasts it would be possible to determine the estimate of the effect in a manner similar to that previously performed.

$$\text{ESTIMATE OF EFFECT}_j = \frac{(\text{contrasts}_j)}{n2^{k-1}} \quad .$$

But the goal is to find the sum of squares associated with each effect. Coincidentally, these sum of squares can also be determined directly from the contrasts with the equation

$$SS_{EFFECT\,j} = \frac{(\text{contrasts}_j)2}{n2^{k}} \quad .$$

Since we only analyzed each parameter at two values, each effect only has one degree of freedom. Therefore, the mean square of the effect is the same as the SS_{EFFECT}.

$$MS_{EFFECT\,j} = \frac{SS_{EFFECT\,j}}{DF} = \frac{SS_{EFFECT\,j}}{1}$$

We are almost ready to find a statistic, but it is first necessary to determine the mean square of the error. The process used to determine the MS_{ERROR} is identical to the one described in Chapter 2. The sum of the squares for the total experiment is simply the sum of the squares of the differences between each measured response and the average of the measured responses at each experimental point,

$$SS_{TOTAL} = \sum_{j=1}^{k} \sum_{i=1}^{n} (y_{ji} - \bar{y}_j)^2 \quad .$$

The sum of squares of the error is then

$$SS_{ERROR} = SS_{TOTAL} - \sum_{j=1}^{k} SS_{EFFECT_j} \quad .$$

The degrees of freedom (DOF) for the error is simply the total number of experiments less the DOF associated with all the effects,

$$DOF = (n-1)\ 2^k.$$

So the mean square of the error is

$$MS_{ERROR} = SS_{ERROR}/((n-1)\ 2^k).$$

The ratio of the mean square of an effect to the mean square of the error gives the F statistic that we use to evaluate the significance of the effect.

To determine if an effect is significant, we test the hypothesis

$$H_0 : \text{Effect} = 0$$

$$H_1 : \text{Effect} \neq 0$$

For φ confidence that an effect is significant, the following must be true,

$$F = MS_{Effect}/MS_{ERROR} \geq F_{(n-1)2^k,2^k,1-\varphi}$$

As stated in the beginning, these experiments are conducted to determine which parameters to include in the study. Since the highest order model one uses in trying to optimize a process is second-order, a resolution V fractional factorial design is adequate for the initial study of the parameters. Yates' algorithm remains valid for fractional factorial designs. The experiments must be arranged in Yates' order and the algorithm is repeated k-p times.

OTHER ISSUES IN CONDUCTING EXPERIMENTS

The basic assumption in conducting designed experiments is that the experimenter can control the parameters that affect the process. Unfortunately, this assumption is not always valid. Subtle differences in equipment, samples, and the performance of the experiments can bias the experimental results. These differences can be as obscure as the effect of the operator learning to use the equipment better as he conducts more experiments, to as obvious as having samples used in the process that were prepared in different ways. These factors are often unavoidable but must be considered. The experimenter has three tools to reduce the factors' influence: replicating, blocking and randomizing. Replicating is simply repeating experiments so that an average response can be used in the analysis. Blocking is the grouping of experiments into smaller sets that are more homogeneous than the entire set. Randomizing is the deliberate effort to avoid patterns in the way one conducts experiments such that present extraneous factors are averaged out of the results.

REPLICATING

Replicating is simply performing the same experiment more than one time. One replicates an experiment to increase confidence in the results. The accuracy of the average of measurements is considered greater than the accuracy of any single measurement. However, one of the goals of using designed experiments is to obtain the most information from the minimum number of experiments. Although replicating improves the results and averages out a number of biases, it defeats this goal. In response surface

studies, one generally replicates experiments to obtain statistical information for the analysis of data and not to average out the effects of biasing elements. For this end, we have blocking and randomizing.

BLOCKING

Blocking is applied when the experimenter identifies factors that could possibly affect the results but are not necessarily parameters of the process. These factors normally set some portion of the set of experiments to be more homogenious than the entire set. Examples include: conducting portions of the experiments on different days; using different vendors for the samples processed in the experiments; or having different operators conduct the experiments. The goal of blocking is to prevent these factors from influencing the factors under investigation. In factorial experiments we block by deliberately confounding the effects of these factors with higher order interactions to keep the lower order interactions more precise. The lower order interactions are the most important in response surface studies.

The procedures used to generate blocks are very similar to those used to generate fractional factorial designs. The most direct approach is to use a relation where the suspect variable is set to equal a higher order interaction. For example, consider a 2^3 factorial design. We designate the extraneous factor as D. Then the relation to generate two blocks is

$$D = ABC.$$

One block is generated when the product ABC is +1 and the other block is generated when the product is - 1. The effect of D is confounded with the interaction of ABC.

Often, more than a single factor needs to be considered in blocks. In such cases, the experimenter must be careful or he may inadvertently confound a lower order interaction. For example, consider the same 2^3 factorial design but with two extraneous factors. Since there are four possible combinations of these extraneous factors -- (−,−), (+,−), (−,+), and (+,+)-- four blocks must be generated. The experimenter can generate four blocks by setting each of the two extraneous factors equal to a different relation of the three experimental parameters. An example of a poor choice is the two equations

$$D = ABC \quad \text{and} \quad E = AB.$$

The blocks generated by the four combinations of D and E are

BLOCK	(D,E)	A	B	C
I	(−,−)	−	+	+
		+	−	+
II	(+, −)	−	+	−
		+	−	−
III	(−,+)	−	−	−
		+	+	−
IV	(+,+)	−	−	+
		+	+	+

Although it is not immediately apparent, a first order effect is confounded with the extraneous factors. The most direct way of seeing this confounding is by evaluating the relation DE. Since

$$D = ABC \quad \text{and} \quad E = AB,$$

$$DE = ABCAB,$$

which is the same as

$$DE = A^2B^2C.$$

Since a factor multiplied by itself is always +1, this reduces to

$$DE = C.$$

DE is confounded with the first order effect C. A better set of equations would be

$$D = AB \quad \text{and} \quad E = AC.$$

Then, only two factor interactions are confounded with the extraneous factors: D with AB, E with AC, and DE with BC.

The following is a step-by-step approach to blocking 2^k factorial experiments.

1. Determine the minimum number of blocks needed. The number of blocks needed is a function of the number of extraneous factors, or the size of the set affected by an extraneous factor. Generally, we accommodate the factors that demand the smallest size blocks. Consider a 2^4 factorial design conducted over a period of two days. Since there are two levels of one extraneous factor, the experiment should be blocked into two sets of eight experiments. One block of eight experimentsis executed on the first day and the remaining blocks executed on the second day. Now suppose we stipulate that only five experiments can be conducted on the first day and as many as 14 on the second day. In this case, the constraint of five on the first day is the governing limitation. The experiment can then be broken down into four blocks of four experiments, with one block of four executed on the first day and the remaining three blocks, for a total of 12 experiments, executed on the second day.

2. Determine the number of extraneous factors needed to accommodate the block size. In this step we determine the minimum number of factors needed to generate the appropriate number of blocks. These factors are abstract and do not exist physically; they are chosen solely for the purpose of generating the appropriate number of blocks and of evaluating the confounding. In the second example above, two factors are needed to generate four blocks, identified as D and E.

3. Determine the generating equations for the blocking that have acceptable confounding. In this step, the experimenter tries to confound the extraneous effects with the higher order interactions in which he is least interested. In the second example of step 1, suppose that we are least interested in the third and higher order interactions. Two possible design generating equations are

$$E = ABC \text{ and } F = ACD.$$

In this set, the extraneous factors are confounded with the ABC, ACD, and BD interactions. If this is unacceptable, we attempt another set of equations. Note that in this example, it is impossible to avoid confounding interactions of less than three factors. The experimenter has to choose which lower order interactions to confound.

4. Identify which experiments belong to each block. In this step, the experimenter simply identifies which experiments satisfy the possible combinations of the design generating factors. Continuing with the second example, there are four possible combinations of E and F: (−,−), (+,−), (−,+), and (+,+). One of these combinations is associated with each block. Consider the first combination, (−,−). and the design generating equations E

= ABC and F = ACD. The four experiments for which the design generating equations yield "- " for both E and F are

A	B	C	D
-	-	-	-
+	-	+	-
+	+	-	+
-	+	+	+.

These four experiments belong to block one.

RANDOMIZING

Randomness is an underlying tenet of statistics. It is assumed in statistics that experimental results are randomly distributed about a population mean. Any pattern that exists in the experiment will violate this assumption and may skew the results. To avoid such a bias, the experimenter must make a deliberate effort to randomly order the experiments. Randomizing is the most apparent way to "average-out" the effects of the experimenter's "learning curve." It is also useful in limiting the effects of some discrete factors that cannot be handled by blocking. But, we do not need to identify a reason to randomize. Factors that we are unaware of can also bias our results. Therefore, randomizing is a standard procedure in conducting experiments.

In conducting designed experiments, the experimenter must attempt to avoid all possible sources of variance. When this is impossible. he has three alternatives: replicating, blocking, or randomizing. Replicating is the most reliable tool but the least desirable since it requires the expenditure of extra resources. Blocking limits the effect of known sources of variance by confounding the variance with higher order interactions that are less critical to the study. Randomizing is the avoidance of patterns in the way one conducts his experiments. As a general rule, the experimenter first attempts to avoid sources of variance. If he cannot, he then attempts to block their effects. For the sources of variance that he cannot block, he applies randomizing with the hope of limiting the effects of these sources of variance.

CHAPTER 4

RESPONSE SURFACES AND OPTIMIZING

In Chapters 2 and 3 we developed the primary tools that allow us to evaluate and make an assessment of what is important in a process. However, our goal is to go one level further and to try to determine the set of conditions that will produce an optimum response. What we need is a method of understanding how the response varies in the space defined by the range of values the parameters can assume and a method of using this knowledge to find the optimum conditions. If the parameters can be varied continuously in their range and we assume that the response changes continuously as the parameters are changed, then the set of all values the response can assume defines a surface within the parameter space. If we can determine the surface's shape, it is reasonable to assume that we could also navigate it and find the optimum condition. Chapter 2 defined the tools needed to define the surface, the least squares regression. All that is necessary is to learn the tools needed to understand and navigate the surface. The three tools that will be presented are: 1) the method of steepest ascent; 2) canonical analysis; and 3) the method of ridge analysis. Before beginning , it is helpful to look at a qualitative illustration of the proposed procedure.

A Qualitative Look at the Response Surface Methodology

Consider a process which has two parameters, a and b, and a true response y that is graphically represented in two dimensions in Figure 4.1. The two axes represent the range of settings for the parameters a and b. The

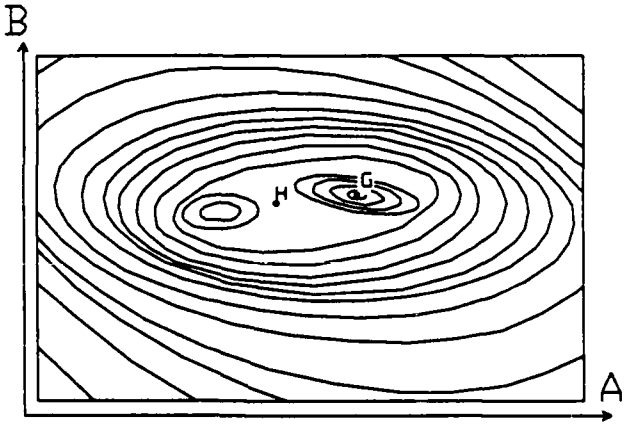


Figure 4.1 The response surface of a two parameter process.

lines represent the contours of constant response. And point G, the maximum response, is considered the optimum condition. Of course, if we knew as much as this graph presents we could immediately choose the values of a and b that would give the response at G. In situations where very little is known of the nature of the response surface, one applies response surface methodology (RSM). RSM, involves strategically selecting experiments that can generate mathematical models that characterize the response surface in discrete regions. Once a model is formed, mathematical methods are used to analyze the characterization to find either the optimum condition or a direction to search for it.

Consider the same process above. If an experimenter knew nothing of the response surface, he would begin by doing experiments. Suppose he began by doing a full factorial experiment at an arbitrary location as Figure 4.2a shows. By applying a least squares regression to this design, he could fit a linear model that reasonably represents the surface, Figure 4.2b. Since an optimum condition does not exist in the modeled region, he would need to

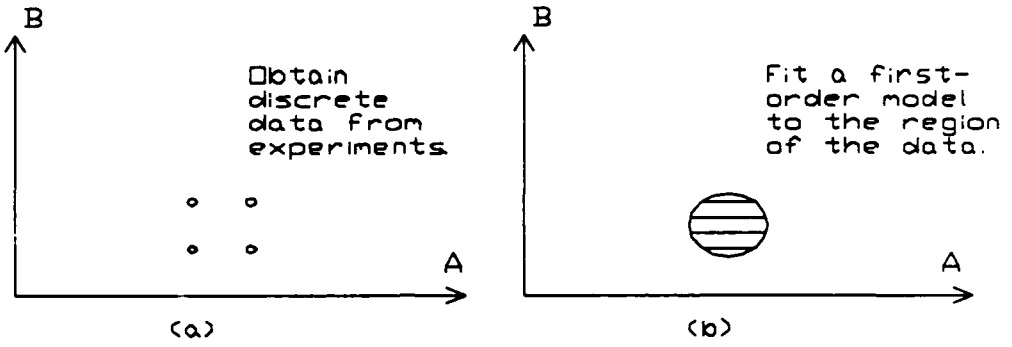


Figure 4.2 a) A full factorial experiment and b) a linear response surface.

analyze the model and determine the direction in which to search for it. The mathematical tool the experimenter could use to find this direction is the method of steepest ascent. The method of steepest ascent is a procedure applied to linear models that determines a direction from the model center ($x_1 = x_2 = 0$) that ascends (or descends if the optimum is a minimum) most directly. The experimenter would then conduct experiments along the direction at fixed intervals until he can gain no further improvement in the response. (See Figure 4.3.)

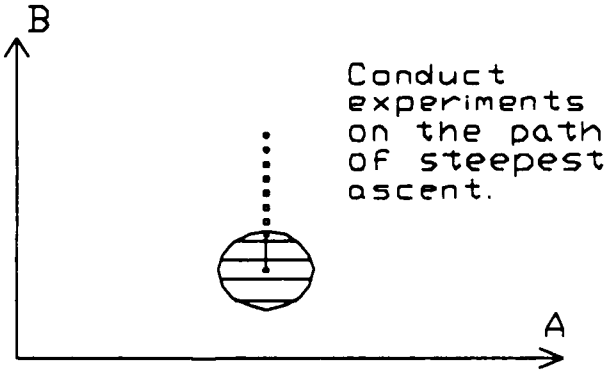


Figure 4.3 The direction of steepest ascent.

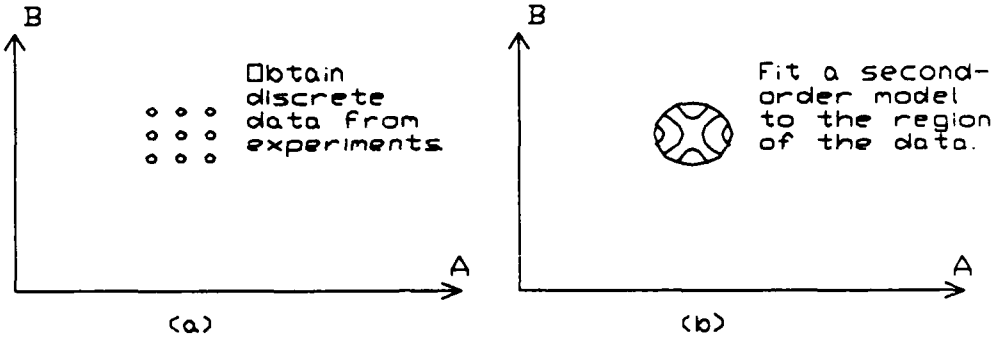


Figure 4.4 a) A central composite experiment and b) a second-order response surface.

Now suppose the experimenter began his experiments in the vicinity of point H on Figure 4.1. In this region, a linear model would not fit the response. Analysis would show the existence of curvature and the necessity for a higher order model. If the experimenter expanded the experiment to a central composite design (Figure 4.4a) he could fit a second-order model to the region (Figure 4.4b).

A second-order model can be one of 7 types: a ridge wth a maximum (Figure 4.5a); a valley with a minimum(Figure 4.5b); a rising ridge (Figure 4.5c); a descending valley (Figure 4.5d); a saddle (Figure 4.5e); or a stationary ridge or valley (Figure 4.5f). None of these would be immediately apparent to the experimenter from the mathematical model. Canonical analysis is the mathematical procedure that would allow the experimenter to determine which type of surface the model represented. This powerful procedure provides valuable information on the nature of the surface and identifies the maximum or minimum points of ridges or valleys. In the example, after canonical analysis, the experimenter would know that the surface was a saddle, and he would know the directions of the ridges and

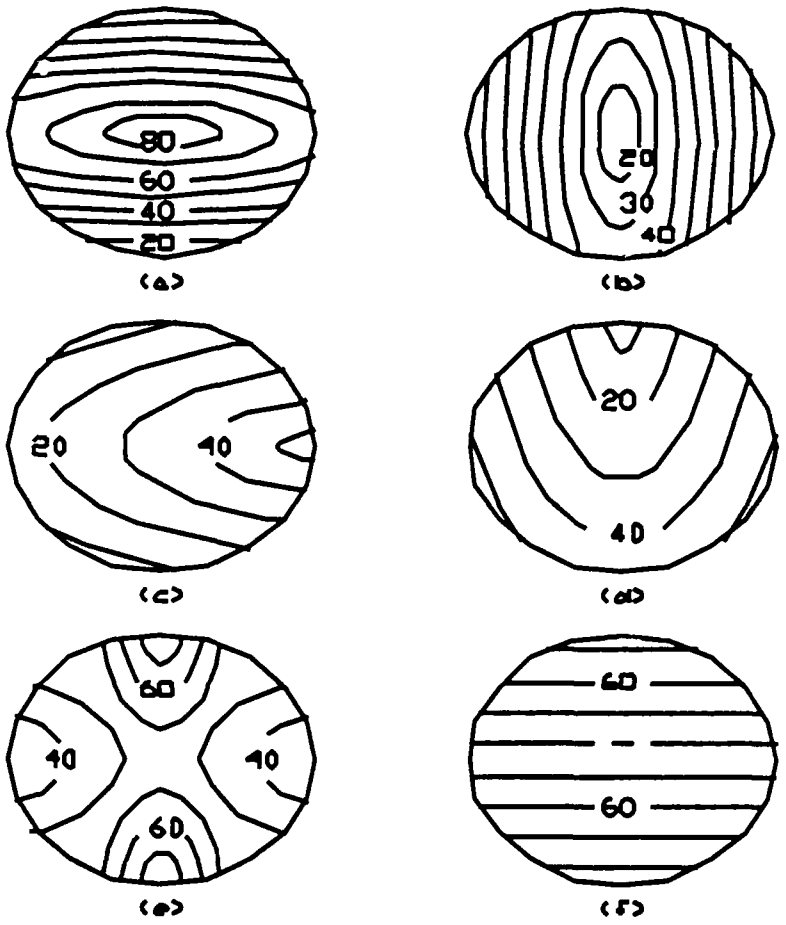


Figure 4.5 Second-order surfaces: a) a ridge with a maximum; b) a valley with a minimum; c) a rising ridge; d) a descending valley; e) a saddle; f) a stationary ridge (or valley).

valleys. Since the optimum condition does not occur in the model, the experimenter would have to conduct further analysis to find a search direction. The mathematical procedure applied to second-order models to find this direction is the method of ridge analysis. This procedure is very similar to the method of steepest ascent, both in theory and in how the

experimenter would apply the results. The method provides a path along which the experimenter should conduct further experiments to find the optimum response.

Combining empirical model building with the three procedures -- the method of steepest ascent; canonical analysis; and the method of ridge analysis -- in an ordered manner allows the experimenter to understand and to navigate the response surface and, in turn, to find the conditions for the optimum response. This chapter explains the three latter procedures and presents a flow diagram for the proper structure of a response surface study.

THE METHOD OF STEEPEST ASCENT

Response surfaces modeled by linear models have no clearly definable maximum or minimum points. Linear models represent planar surfaces which ascend or descend in the same direction from all points on the surface. These directions are perpendicular to the contours of constant response (Figure 4.6). One can determine these directions by the application

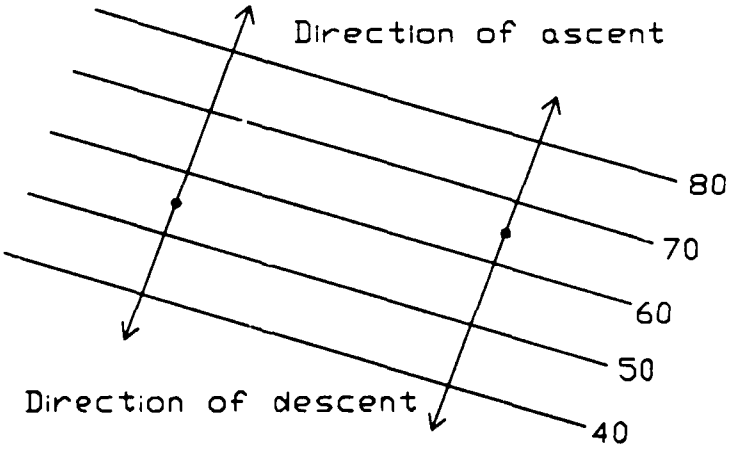


Figure 4.6 The direction of steepest ascent for the linear model

of a Lagrangian multiplier for restricted optimization. Lagrange's observation is that if an x_0 maximizes (or minimizes) $F(x)$ subject to the side condition $Q(x) = 0$, then $\nabla F(x_0)$ and $\nabla Q(x_0)$ are parallel. Thus, if $\nabla Q(x_0) \neq 0$, then a scalar μ exists such that

$$\nabla F(x_0) = \mu \nabla Q(x_0).$$

μ is the Lagrangian multiplier.

In the method of steepest ascent, the Lagrangian multiplier is applied to find the direction of ascent (or descent) from the experiment center

$$[x_1, x_2, \ldots, x_k] = (0, 0, \ldots, 0).$$

This is accomplished by evaluating all points equidistant from the design center using the relation

$$\sum_{i=1}^{k} x_i^2 = R^2.$$

From this relation one forms $Q(x) = 0$,

$$Q(x_1, x_2, \ldots, x_k) = \sum_{i=1}^{k} x_i^2 - R^2. \quad [1]$$

$F(x)$ is already known to be,

$$F(x_1, x_2, \ldots, x_k) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i.$$

1 The equation $Q(x) = 0$ defines a surface that encloses a portion of the parameter space. Qualitatively, we are attempting to find where on this surface the function $F(x)$ is a maximum (or minimum). The $Q(x)$ shown here defines the surface equidistant from the center of the design. By incrementing R we can find the maximum (or minimum) of $F(x)$ at different distances from the design center. Together, these points define the path of steepest ascent (or descent). When $F(x)$ is a linear model, the identification of these paths is greatly simplified since all points of the path fall on a straight line.

Taking the gradient of both equations,

$$\frac{\partial Q(x_1,x_2,\ldots,x_k)}{\partial x_i} = 2\, x_i \qquad (i = 1, 2, \ldots, k),$$

and

$$\frac{\partial F(x_1,x_2,\ldots,x_k)}{\partial x_i} = \beta_i \qquad (i = 1, 2, \ldots, k),$$

and applying the Lagrangian multiplier, as stated above, it is possible to form the series of relations

$$\beta_i = 2\, \mu\, x_i \qquad (i = 1, 2, \ldots, k)$$

or

$$x_i = \beta_i / (2\, \mu) \qquad (i = 1, 2, \ldots, k).$$

These simple equations define the points on the path of steepest ascent or descent. The vector of all x_i from i = 1 to k defines the direction. So, to determine the increment by which to vary the parameters on the steepest path, one need only to select a convenient value of μ. For example, consider the three factor design, used in Chapter 2, with the parameter values

	-1	1
a	10	20
b	90	150
c	65	95 .

Let us say the linear model has the coefficients

$$\beta_0 = 15$$

$$\beta_1 = 4$$

$$\beta_2 = 7.2$$

$$\beta_3 = -8.5 \ .$$

Suppose also that factor a can only be adjusted by whole units. That is one can only vary a by n, an integer, from the uncoded design center; meanwhile, b and c can be varied continuously. In this case, it is necessary to choose a μ corresponding to the uncoded increment of 1 unit of a or .2 as the coded variable. The μ for the path becomes

$$\mu = \beta_i/2x_i = 4/2(.2) = 10 .$$

So the increments of the coded variables are

$$x_1 = .2 ,$$
$$x_2 = 7.2/2(10) = .36,$$
$$x_3 = -8.5/2(10) = -.425,$$

which correspond to the uncoded values of $x_1 = 1$, $x_2 = 10.6$, and $x_3 = -6.375$. The paths of steepest ascent and descent are

ASCENT	DESCENT
15, 110.0, 80.00	15, 110.0, 80.00
16, 120.6, 73.63	14, 99.4, 86.38
17, 131.2, 67.25	13, 88.8, 92.75
18, 141.8, 60.88	12, 78.2, 99.13
19, 152.4, 54.50	11, 67.6, 105.50
20, 163.0, 48.13	10, 57.0, 111.88
21, 173.6, 41.75	21, 46.4, 118.25 .

This procedure provides a simple method for determining the path of steepest ascent or descent. It is important to note, however, that the manner

in which we code the variables will affect the direction. Consider the same model modified such that the uncoded variables are used. The model equation becomes

$$y = 15 + 4\frac{(A-15)}{5} + 7.2\frac{(B-110)}{30} - 8.5\frac{(C-80)}{15}.$$

The equation can be reduced to

$$y = 21.93 + .8A + .24B - .567C.$$

By applying the restriction

$$(A - 15) + (B - 110) + (C - 80) = R$$

the optimizing equations become

$$(A - 15) = .8/2\mu\,,\ (B - 110) = .24/2\mu\,,\ \text{and}\ (C - 80) = -.567/2\mu\ .$$

The increments of the uncoded variables are

$$A = 1,$$
$$B = .24/2(.4) = .3,$$
$$C = -.567/2(.4) = -.708\ ,$$

and the paths of steepest ascent and descent are

ASCENT	DESCENT
15, 110.0, 80.00	15, 110.0, 80.00
16, 110.3, 79.29	14, 109.7, 80.71
17, 110.6, 78.58	13, 109.4, 81.42
18, 110.9, 77.88	12, 109.1, 82.13

ASCENT	DESCENT
19, 111.2, 77.17	11, 108.8, 82.83
20, 111.5, 76.46	10, 108.5, 83.54
21, 111.8, 75.75	21, 108.2, 84.25 .

Although there is a mathematical basis for the path of steepest ascent the experimenter must still make a decision as to which path to follow. As a general rule, the experimenter should put more credence in the path generated by the coded variables.

THE CANONICAL ANALYSIS OF SECOND-ORDER FITTED SURFACES

Chapter 2 stated without explanation that the second-order model was the highest order model usually used in the analysis of response surfaces. It is at this point that we begin to explore the reasons for this statement. In the analysis of curved response surfaces, the experimenter wants to determine several items of information dealing with the nature of the surface and the location of the conditions for an optimum response. These include:

- whether an optimum response occurs in the vicinity of the experiment,
- if so, whether the optimum response occurs at a point or over a region,
- if not, where the optimum response is most likely to be found.

One possible approach that is not model specific is to take the derivative of the response with respect to the vector of parameters, **x**, and to evaluate it at

zero, $\delta y/\delta \mathbf{x} = 0$. The application of this method to the second order model is shown below

$$\delta y/\delta \mathbf{x} = \delta[\mathbf{x}'\mathbf{b} + \mathbf{x}'B\mathbf{x}] / \delta \mathbf{x} = 0$$

$$0 = \mathbf{b} + 2B\mathbf{x}.$$

The vector of parameters that solves this equation is

$$\mathbf{x_0} = -B^{-1}\mathbf{b}/2.$$

This vector represents a single point on the response surface that is either a *maximum*, a *minimum*, or a *saddle point*. The actual nature is not readily determined by differential calculus. In addition, its location may be far from the design center, which prevents making any conclusions about it at all. Fortunately, the second-order model provides the opportunity to apply a different approach to the analysis of response surfaces. This approach uses the tools of linear algebra and matrices.

The second-order model has three terms, b_0, $\mathbf{x}'\mathbf{b}$, and $\mathbf{x}'B\mathbf{x}$. Of these terms, only the third term contributes to the curvature of the surface. This implies that the nature of the surface can be inferred from the matrix B. Three facts are known about the matrix B: it is real; it is symmetric; and its dimension is k. From these facts and from the theory of matrices, we know further that the matrix B has k eigenvalues and k linearly independent eigenvectors that may be chosen to be orthonormal. Recall the relation between eigenvalues, λ, and the eigenvectors, $\mathbf{m}$, for a matrix B,

$$B\mathbf{m} = \lambda\mathbf{m}.$$

The equation states that along certain directions defined by the eigenvectors the transformation caused by the matrix B can be reduced to a scalar

operation by the corresponding eigenvalue. That is, along these directions a uniform curvature exists. Such conditions occur only at ridges and valleys of second-order surfaces. Hence, by knowing the sign and magnitude of the eigenvalues of B, one can know the nature of the surface defined by the model. Further information can be obtained, such as the exact locations of the features, by transforming the second-order model to a form that uses the orthonormal eigenvectors of B as the new axes. This simplified version of the model equation is called its canonical form. Further, using the notation of Box and Draper in [1], if the axes of the model are centered at the design center, the model is in A canonical form and if the axes of the model are centered at the system center (the stationary point identified by differentiation) it is in B canonical form.[2]

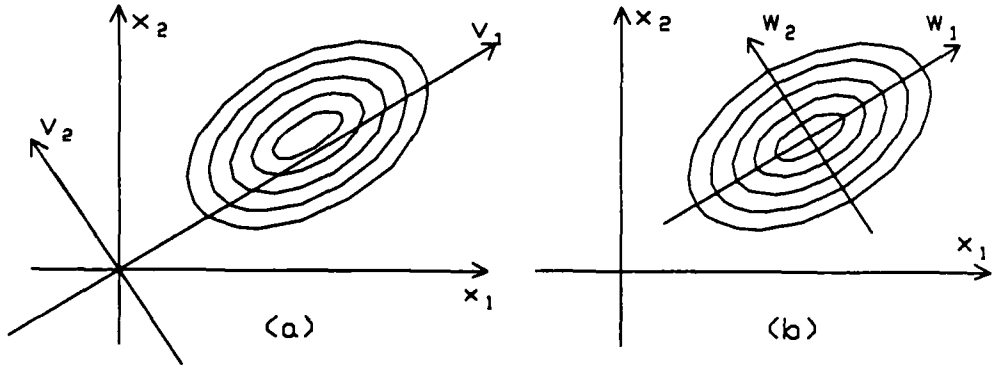


Figure 4.7 The canonical axes of the a) A canonical form and the b) B canonical form.

2 The design center is the center of the experiment where x1=x2=...xk=0. The system center is the stationary point of the model defined by $\mathbf{x_0}=-\mathbf{B}^{-1}\mathbf{b}/2$.

The A canonical form of the model looks like

$$\hat{y} = b_0 + \theta_1 v_1 + \ldots + \theta_k v_k + \lambda_1 v_1^2 + \ldots + \lambda_k v_k^2.$$

This form is obtained by using the conversion of coordinates,

$$\mathbf{x} = M'\mathbf{v}$$

where M is the matrix formed by using the normalized eigenvectors as its columns. Applying the conversion to the second-order model the equation of the second-order model becomes

$$\hat{y} = b_0 + (\mathbf{x}'M)(M'\mathbf{b}) + (\mathbf{x}'M)(M'BM)(M'\mathbf{x})$$

Recalling that B may be written as M'LM, where L is the diagonal matrix of *eigenvalues with* λ_i *corresponding to the eigenvector in column* i of M and that MM' = I, the equation becomes

$$\hat{y} = b_0 + \mathbf{v}(M'\mathbf{b}) + \mathbf{v}'L\mathbf{v} \ .$$

So by writing $\theta = (M'\mathbf{b})$ we obtain the desired form.

The B canonical form of the model looks like

$$\hat{y} = b_0 + \lambda_1 w_1^2 + \ldots + \lambda_k w_k^2.$$

This form is obtained by using the conversion of coordinates

$$(\mathbf{x} - \mathbf{x_0}) = M'\mathbf{w},$$

where M is the same matrix as before and y_0 is the response at the stationary point. To show that this conversion is valid let us apply a different procedure. First, let us determine y_0 by evaluating the model at $\mathbf{x_0}$

$$y_0 = b_0 + \mathbf{x_0}'\mathbf{b} + \mathbf{x_0}'B\mathbf{x_0} \ .$$

This can be reduced to terms of b_0, $\mathbf{x_0}$, and $\mathbf{b}$. Substituting $\mathbf{x_0} = -\mathbf{B}^{-1}\mathbf{b}/2$ into this equation

$$\begin{aligned} y_0 &= b_0 + [-\mathbf{B}^{-1}\mathbf{b}/2]'\mathbf{b} + [-\mathbf{B}^{-1}\mathbf{b}/2]'\mathbf{B}[-\mathbf{B}^{-1}\mathbf{b}/2] \\ &= b_0 - \mathbf{b}'\mathbf{B}^{-1}\mathbf{b}/2 + \mathbf{b}'\mathbf{B}^{-1}\mathbf{B}\mathbf{B}^{-1}\mathbf{b}/4 \\ &= b_0 - \mathbf{b}'\mathbf{B}^{-1}\mathbf{b}/2 + \mathbf{b}'\mathbf{B}^{-1}\mathbf{b}/4 \quad . \end{aligned}$$

Now, by reinserting $x_0 = -\mathbf{B}^{-1}\mathbf{b}/2$, the equation becomes

$$\begin{aligned} y_0 &= b_0 + \mathbf{x_0}'\mathbf{b} - \mathbf{x_0}'\mathbf{b}/2 \\ &= b_0 + \mathbf{x_0}'\mathbf{b}/2 \quad . \end{aligned}$$

Second it is necessary to translate the axes to the new origin at $\mathbf{x_0}$. To do this, we define the vector $\mathbf{z}$ such that $\mathbf{z} = (\mathbf{x} - \mathbf{x_0})$ and then write the response function in terms of $\mathbf{z}$,

$$\begin{aligned} \hat{y} &= b_0 + (\mathbf{z}' + \mathbf{x_0}')\mathbf{b} + (\mathbf{z}' + \mathbf{x_0}')\mathbf{B}(\mathbf{z} + \mathbf{x_0}) \\ &= b_0 + \mathbf{x_0}'\mathbf{b} + \mathbf{x_0}'\mathbf{B}\mathbf{x_0} + \mathbf{z}'\mathbf{b} + \mathbf{z}'\mathbf{B}\mathbf{x_0} + \mathbf{x_0}'\mathbf{B}\mathbf{z} + \mathbf{z}'\mathbf{B}\mathbf{z}. \end{aligned}$$

Since $\mathbf{z}^{-1}\mathbf{B}\mathbf{x_0}$ is equivalent to $\mathbf{x_0}'\mathbf{B}\mathbf{z}$,

$$\hat{y} = b_0 + \mathbf{x_0}'\mathbf{b} + \mathbf{x_0}'\mathbf{B}\mathbf{x_0} + \mathbf{z}'(\mathbf{b} + 2\mathbf{B}\mathbf{x_0}) + \mathbf{z}'\mathbf{B}\mathbf{z}.$$

Since the first three terms are the same as the stationary point, it is possible to substitute y_0 for $b_0 + \mathbf{x_0}'\mathbf{b} + \mathbf{x_0}'\mathbf{B}\mathbf{x_0}$. So,

$$\hat{y} = y_0 + \mathbf{z}'(\mathbf{b} + 2\mathbf{B}\mathbf{x_0}) + \mathbf{z}'\mathbf{b}\mathbf{z}.$$

Substituting $-\mathbf{B}^{-1}\mathbf{b}/2$ for $\mathbf{x_0}$, yields

$$\begin{aligned} \hat{y} &= y_0 + \mathbf{z}'(\mathbf{b} + 2\mathbf{B}(-\mathbf{B}^{-1}\mathbf{b}/2)) + \mathbf{z}'\mathbf{B}\mathbf{z} \\ \hat{y} &= y_0 + \mathbf{z}'\mathbf{B}\mathbf{z} \quad . \end{aligned}$$

Since we have $\mathbf{z} = M'\mathbf{w}$ as our transformation, we know

$$\begin{aligned}\mathbf{z}'B\mathbf{z} &= \mathbf{w}'MBM'\mathbf{w}, \\ &= \mathbf{w}'MM'LMM'\mathbf{w}, \\ &= \mathbf{w}'L\mathbf{w},\end{aligned}$$

and our model becomes

$$\hat{y} = y_0 + \mathbf{w}'L\mathbf{w},$$

which is the desired form.

In canonical analysis, the type of canonical form used is determined by the location of the stationary point. If the stationary point is close to the experiment center, the B canonical form is used. In such a situation, the axes of the model define the regions and directions of the near constant, or the most rapidly increasing or decreasing responses. With this information the experimenter can determine whether the optimum condition is present, over what regions he can change the parameters without affecting the response, or in what direction he must conduct more experiments to search for the optimum response. If the stationary point is outside the experimental space, the information of the axes of the B canonical form is not always valid. In this case, rather than navigating the ridges, the experimenter searches for the regions of constant response or the directions to the optimum response. The A canonical form is the most useful for this analysis.

The analysis of a model in the B canonical form is most obvious. If the eigenvalues (the model coefficients) are all negative, the stationary point is a maximum. Conversely, if the eigenvalues are all positive, the stationary point

is a minimum. If the eigenvalues are combinations of positive and negative values, the stationary point is a saddle. The ridge identified by the most negative eigenvalue is the direction of the most rapidly decreasing response. Conversely, the ridge identified by the most positive eigenvalue is the direction of the most rapidly increasing response. The ridges identified by eigenvalues close to 0 are regions of near constant response.

The analysis of a model in the A canonical form is quite different. The analysis involves the comparison of four sets of values defined as the slopes, the curvatures, the axial distances of the system origin from the design center, and the changes in the response along each axis. The slopes are simply the coefficients of the linear portion of the model,

$$\Theta = (\mathbf{M'b}).$$

The curvatures are the coefficients of the second order portion or, more simply, the eigenvalues λ_i. The axial distances of the design center from the stationary point are simply the negative values of the stationary point in the new system of coordinates. By evaluating the derivative of the A canonical model the distances can be shown to be

$$V_{i0} = -\theta_i/(2\lambda_i) \quad (i = 1, 2, \ldots, k).$$

The changes in y are evaluated by comparing the responses at the model's origin with the responses along each axis at a distance of $\pm$ 1 units. So for the i^{th} term three responses are evaluated on the i^{th} axis,

Response	Coordinates $(v_1, \ldots, v_i, \ldots, v_k)$
$\hat{y}_c$	$(0, \ldots, 0)$
$\hat{y}_c + \theta_i + \lambda_i$	$(0, \ldots, 1, \ldots, 0)$
$\hat{y}_c + \theta_i + \lambda_i$	$(0, \ldots, -1, \ldots, 0)$.

The average response is $y_c + (2/3)\lambda_i$ with a sample variance of $\theta_i^2 + (1/3)\lambda_i^2$. The term used to measure the change in y is the range of this evaluation. In statistics, the standard deviation is approximated by (sample range)/$3^{1/2}$. Therefore, the sample range for the i^{th} element is

$$r_i = (3\theta_i^2 + \lambda_i^2)^{1/2}.$$

Combining these values in a table, yields

AXIS	1	2 ...	k
SLOPES	θ_1	θ_2 ...	θ_k
CURVATURES	λ_1	λ_2 ...	λ_k
DISTANCE	V_{10}	V_{20} ...	V_{k0}
CHANGES	r_1	r_2 ...	r_k

With the table it is possible to evaluate the surfaces in the experimental region. The slopes and curvatures give an indication of which terms, linear or quadratic, dominate in a particular axis direction. As before, the curvatures identify the surface as either a ridge, a valley, or a saddle. The V_{i0}'s are useful in determining the closest distances to the ridges and valleys. Finally, the

ranges give a measure of how fast the response changes along the axes. Exploitation of this information involves identifying the nature of each axis of the system (not the design) and assessing whether these axes are close to the experiment center.

As a general rule, the experimenter first analyzes the table for regions of near constant response that are within the experimental space and that represent the type of region, relative maxima or minima, considered optimum. If these exist, then in most cases the experimenter has identified the optimum conditions. If the regions of near constant response do not exist or if they do not meet the optimum criteria, the experimenter searches for ridges and valleys in the experimental space that rise or fall most rapidly. If these are the dominant ridges of their type in the system, then the experimenter can proceed along these features to search for the optimum conditions.

The regions of near constant response are identified by curvatures close to 0. The closest distance, **RD**, to a point on the region from the experiment's center is identified by the geometric mean of all the axes' distances other than those asociated with the small curvature,

$$\mathbf{RD}^2 = \sum_{i=1}^{k} \begin{cases} V_{i0}^2 & \text{if } \lambda_i \neq 0 \\ 0 & \text{if } \lambda_i = 0 \end{cases}.$$

If the distance is within the experimental space, then it is reasonable to proceed to the region. The most direct manner to reach the region is to progress V_{i0} units along each axis with nonzero curvature. Once at this point,

the experimenter can move along the axis of near constant response by using the axis transformation of the B canonical form:

$$(\mathbf{x}-\mathbf{x}_0) = \mathbf{M}'\mathbf{w}.$$

Of course, we would move to these regions only if they were of the optimum type. For a maximum stationary region, all the nonzero curvatures must be negative. For a minimum stationary region, all the nonzero curvatures must be positive. If both negative and positive nonzero curvatures are present, a "saddle" region exists and the experimenter must proceed to the second method of evaluating the surface.

If no regions of constant response exist, or if the regions that do exist are outside the experimental space or do not represent the optimum condition, then the second method of analysis is used. In this case, the ridge or valley of the system that most rapidly approaches the optimum condition is identified. If the system is not a saddle and a maximum is desired, the most negative curvature identifies the feature; conversely, if the system is not a saddle and a minimum is desired, the most positive curvature identifies the feature. If the system is a saddle the most positive curvature identifies the feature leading to the maximum response and the most negative curvature identifies the feature leading to the minimum response. The closest distance to these features, **AD**, is determined in the same manner as that used to determine the distance to the regions of near constant response. The distance is the geometric mean of all the V_{i0}'s less the one associated with the feature,

$$AD_j = \left(\sum_{i=1}^{j-1} V_{i0}^2 + \sum_{i=j+1}^{k} V_{i0}^2 \right)^{1/2} \quad j = \text{feature number.}$$

If the distance is within the experimental space, then the response surface implies that the optimum condition can be found by proceeding along this axis. The experimenter then moves to the feature by moving along each axis of the canonical form V_{i0} units except for the axis associated with the feature. The experimenter can then move along the feature by using the transformation of the B canonical form:

$$(\mathbf{x}-\mathbf{x}_0) = \mathbf{M}'\mathbf{w}.$$

As in the method of steepest ascent, the information gives a direction for further experiments.

Situations exist where no region or axis of interest can be found within the experimental space, or where the axes that are present represent the opposite type feature (a relative minima when a maximum is desired or vice versa). In these situations, these methods are not applicable. A rough approach is to conduct experiments along the axis with the greatest r value. This approach, however, will not lead the experimenter to the optimum point in the most direct manner. For reaching this point, a third procedure is presented. It is the method of ridge analysis.

THE METHOD OF RIDGE ANALYSIS

The method of ridge analysis is identical in principle to the method of steepest ascent. However, the method of ridge analysis differs from the method of steepest ascent in two ways: First, the experimenter applies the Lagrangian multiplier for restricted maximization to the second-order model

rather than to the linear model; and, second, the experimenter applies the procedure repeatedly for each point on the path, since a path, not a direction, is sought. Since the path extends from the center of the experiment, it is the path of ascent or descent in which the experimenter can have the most confidence.[3]

In the method of ridge analysis, we use the same boundary condition as used in the method of steepest ascent,

$$Q(\mathbf{x}) = \mathbf{x}'\mathbf{x} - R = 0,$$

and apply it to the second-order model,

$$\hat{y} = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'B\mathbf{x},$$

to get the equation

$$F = \quad y - \mu(\mathbf{x}'\mathbf{x} - R^2).$$

Setting the derivative $dF/d\mathbf{x} = 0$, we have

$$dF/d\mathbf{x} = b + 2B\mathbf{x} - 2\mu\mathbf{x} = 0,$$

which can be reduced to

$$(B - \mu I)\mathbf{x} = -\mathbf{b}/2. \qquad (**)$$

3 Since a rotatable experiment is used to generate the model, the variance of the predicted response is a function of distance from the center of the experiment. Given two points with identical predicted responses , one can have more confidence in the response of the point closest to the center of the experimental design.

As in the method of steepest ascent, one can determine the stationary points on the boundary Q(**x**) by selecting different values of μ, but, unlike the method of steepest ascent, all values of μ do not define points on the same path. A boundary of a k parameter quadratic model may have up to 2k stationary points. This can be seen directly from equation (**) above. As μ approaches the values of the characteristic roots of B, the vector **x** increases. Mapping the distance R generated from the vector **x** (**x'x** = R) shows that R approaches infinity as μ approaches each characteristic root. Figure 4.8 shows a situation where a four parameter model with four characteristic roots (0.449, -1.647, -2.025, -2.684) can have 8 stationary points on the same boundary of radius R. In the example, this occurs at any radius greater than 6.5. Of course, these points are not on the same path. Additionally, their responses are not the same. So the question that naturally follows is which

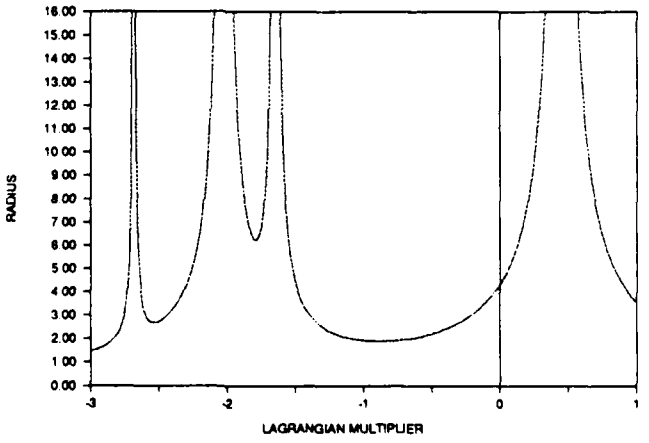


Figure 4.8 The plot of the radius versus the Lagrangian multiplier

point is on the path we seek. Three rules, which are presented without proof, help us make this determination.[4]

1. The trace of stationary points as R increases represent either the absolute maximum, the absolute minimum, a local maximum, or a local minimum. On these features, the response y changes in one of four ways as R increases:

   a. Increases monotonically.

   b. Decreases monotonically.

   c. Passes through a maximum and then decrease monotonically.

   d. Passes through a minimum and then increase monotonically.

2. If $R_1 = R_2$ and $\mu_2 > \mu_1$ then $\hat{y}_2 > \hat{y}_1$.

3. If $\mu_1 > \lambda_i$ for all i, then the stationary point $\mathbf{x}_1$ on the sphere R_1 is a point where the response, $\hat{y}$, attains a local maximum; and, if $\mu_1 < \lambda_i$ for all i, then the stationary point $\mathbf{x}_1$ on the sphere R_1 is a point where the response $\hat{y}_1$ attains a local minimum.

In the effort to optimize processes, rule three is the most useful. From the rule we know that if we select μ greater than the largest characteristic root of B and allow it to approach the largest root, we can determine the path of steepest ascent. And, if we select μ less than the smallest characteristic root of B and allow it to approach the smallest root, we can determine the path of steepest descent. But from rule one we also know that the path may pass through a maximum or minimum and then change direction. Therefore, the

4 These rules were adapted from a list of possible results listed in [1:378].

standard method of analyzing these paths is graphical. The procedure has three steps.

1. Determine the stationary points as m is changed and then calculate the response $\hat{y}$ and the radius R associated with each of these points.

2. Make two plots. On the first, plot the responses versus the radius R. On the second, plot each of the parameter values versus the radius R.

3. On the first plot, a maximum or minimum can be readily identified. By relating the R of this point to the second graph, the experimenter can determine the parameter settings for its condition. If there is no maximum or minimum, the second plot can be used to determine the points at specific intervals of R for further experiments.

Consider the model represented by the equation

$$\hat{y} = [.1088] + \mathbf{x}'\begin{bmatrix} .0247 \\ -.0045 \\ -.0957 \\ .0089 \end{bmatrix} + \mathbf{x}'\begin{bmatrix} .0670 & -.0058 & -.0039 & -.0048 \\ -.0058 & .2195 & -.0245 & -.0173 \\ -.0039 & -.0245 & .0770 & -.0036 \\ -.0048 & -.0173 & -.0036 & .0620 \end{bmatrix}\mathbf{x} .$$

The B matrix of this model has the characteristic roots

$$\lambda_1 = .2254 \qquad \lambda_2 = .0763 \qquad \lambda_3 = .0697 \qquad \lambda_4 = .0540$$

If a minimum is desired, μ is selected initially to be less than λ_4 and then allowed to approach λ_4. If a maximum is desired, μ is selected to be greater than λ_1 and then allowed to approach λ_1. Figures 4.9a and b are the plots when a minimum is desired. As seen on Figure 4.9a, the path proceeds to a

minimum and then starts to increase. This minimum occurs at an R =0.66.

By using Figure 4.9b we can find the parameter settings at this R,

$x_1 = -0.142$

$x_2 = 0.076$

$x_3 = 0.638$

$x_4 = -0.025$.

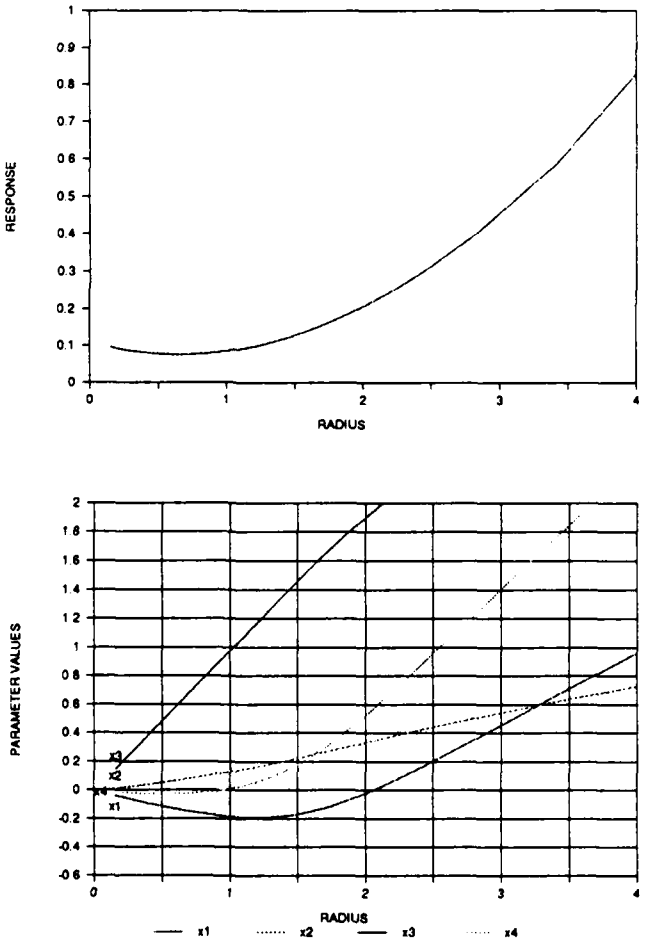


Figure 4.9 Method of ridge analysis applied to obtain a minimum a) response vs. radius and b) parameter settings vs. radius.

These are the predicted settings for the optimum condition. In this case, they are also the coordinates of the center point of the system, the same as those determined in the canonical analysis. The point is that with this system, if a minimum is desired, the study could have ended with canonical analysis. The opposite is true, if a maximumis desired. Since the B matrix has four negative eigenvalues, the system is a valley. Canonical analysis does not provide a strategy for searching for a maximum when the system is a valley. The axes of a valley represent relative minima and are the incorrect features on which to search for a maximum. Therefore, we would have to use the method of ridge analysis. Figures 4.10a and b are the plots of the ridge analysis when a maximum is desired. As we would expect for a valley, Figure 4.10a shows no clear maximum condition. Therefore, Figure 4.10b takes on a new level of importance: It defines the path of steepest ascent. From this graph we can extrapolate parameter settings as we increment the distance R. These points are along the path of steepest ascent and represent uniform increments along which further experiments would be conducted.

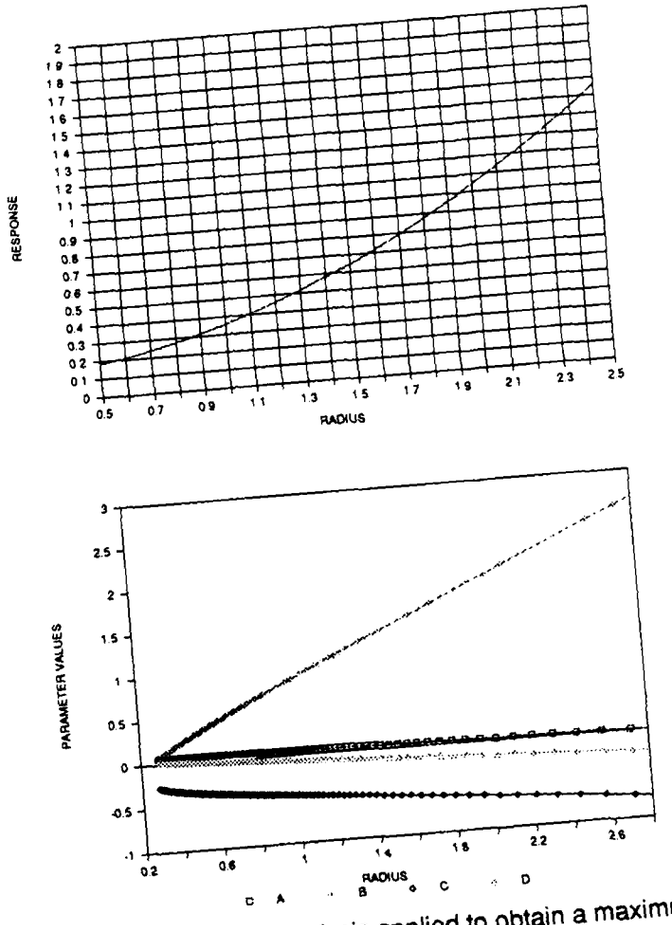


Figure 4.10 The method of ridge analysis applied to obtain a maximum a) response vs radius, and b) parameter settings vs radius.

WORK FLOW FOR OPTIMIZING A PROCESS

The optimization of a process can be separated into three distinct levels. The first-level study is the initial investigation of the process when the experimenter determines which parameters and responses are significant. In the second level of study, the experimenter uses first-order techniques to

study the response surface. Finally, in the third level of study, the experimenter uses second-order techniques. It is possible to find the conditions for an optimum response from any one of these levels of study. The advancement from level to level does not necessarily mean that the experimenter is closer to his goal, but means that the effort and complexity of the study increases. So it is always best to study a process at the lowest level possible and advance to a higher level only when the methods of the lower level become inadequate.

At the first level of study the experimenter identifies the problem and attempts to reduce it to its simplest definition. That is, he identifies which responses are important, defines the objective for the responses, and identifies which parameters of the process should be varied to achieve these objectives. Often, it is not well understood which parameters are significant; so, experiments are conducted to study them. Factorial experiments, as presented in Chapter 3, are used in these studies. Once the problem is reduced to its simplest level, the experimenter then attempts to evaluate it with existing models. If this is not possible, or if the model is unsuccessful at predicting the optimum response, the experiment proceeds to a second-level study. Figure 4.11 is the flow diagram for this level of study.

At the second level, the experimenter conducts experiments to build a first-order model. The experimenter can build the model by conducting a suitable full or fractional factorial experimental design and applying the method of least squares regression. If the model has a serious lack of fit, the experimenter attempts various transformations of the parameters and the response to improve the fit. If the fit cannot be improved, or if evidence of

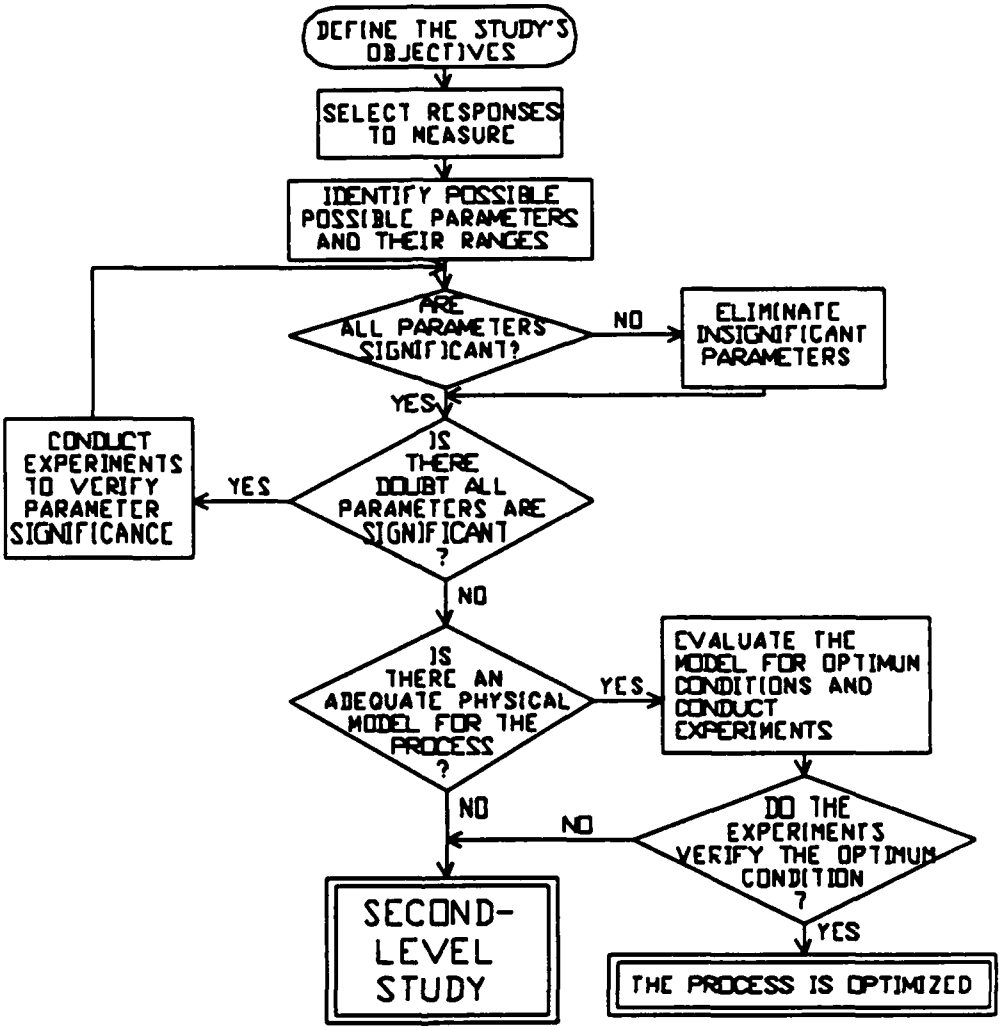


Figure 4.11 The flow diagram of the first-level study.

curvature exists, the experimenter proceeds to a third-level study.

Conversely, if the data generates an adequate model, the experimenter uses it to explore for improved responses. The method of steepest ascent provides the direction of exploration. The experimenter conducts experiments along this direction until he fails to obtain an improved response or reaches a

parameter's limit. If the response moves away from the optimum, the procedure starts over. If all the parameters reach their limit before a change occurs in the direction of the response, then the process is considered optimized. Figure 4.12 is the flow diagram of this level of study.

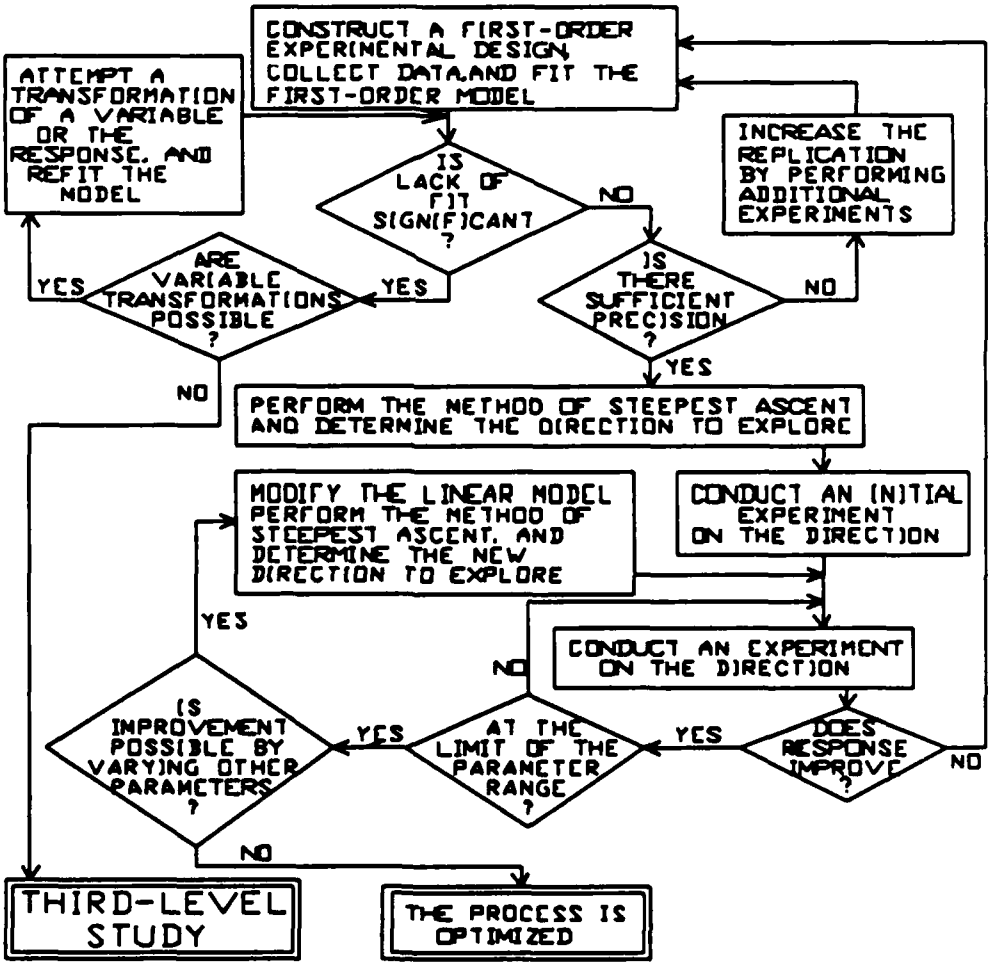


Figure 4.12 The flow diagram of the second-level study.

At the third level the experimenter conducts experiments to build a second-order model. The experimenter can build the model by conducting a central composite designed experiment and applying the method of least squares to its data. As in the second-level study, if there is a significant lack of fit, transformations of the variables and the response are attempted to improve the model. At the end of this iterative cycle, the model with the best fit is used even if the lack of fit is significant. The model is first analyzed using canonical analysis. If canonical analysis reveals that an optimum condition exists, the experimenter verifies the prediction with appropriate experiments. If an optimum condition does not exist, the experimenter attempts to find a path along which the optimum is most likely to be found. This path is either identified using the information gained from the canonical analysis or by applying the method of ridge analysis. Experiments are conducted iteratively along the path until no improvement can be gained. At that point the experimenter makes the decision to continue to conduct experiments or to end the study. Figure 4.13 shows a flow diagram for this level.

The flow diagrams in Figures 4.11 to 4.13 by no means exhaust the possible options available to the experimenter studying response surfaces. As the experimenter gains knowledge of the process from his studies, he may identify different parameters or responses that must be considered. Additionally, other factors not even considered in this discussion may influence the course of the study such as changes or modifications in the equipment. The final course that a response surface study takes is highly dependent on the experimenter's judgement.

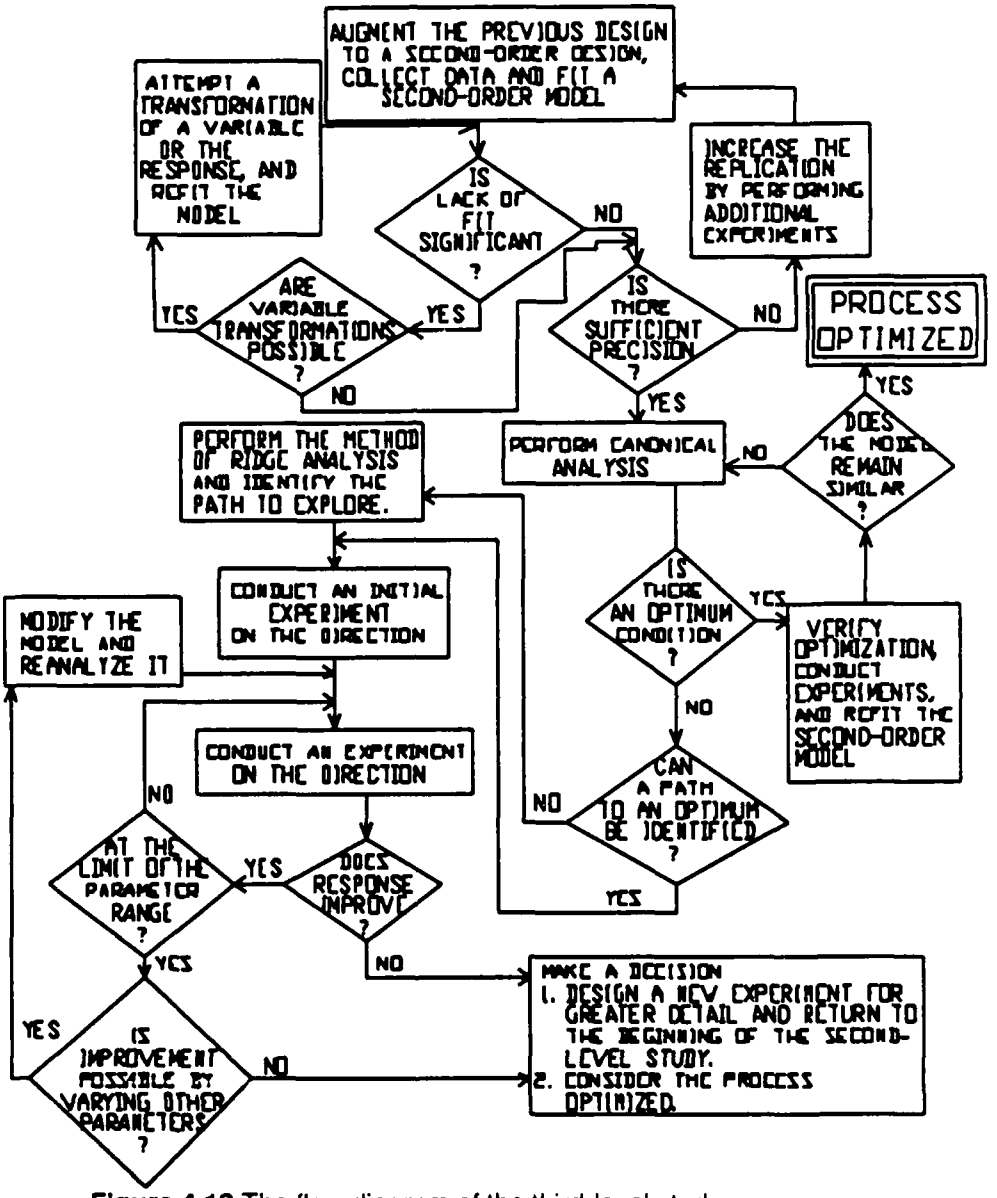


Figure 4.13 The flow diagram of the third-level study.

OPTIMIZING MULTIPLE RESPONSES

In many processes multiple responses must be optimized simultaneously. Several strategies can be applied to this problem, but none are perfect. The following are examples. At level one, if it is determined that two responses approach their optimum in opposite directions, the experimenter can use a single response that is a ratio of the two as the single condition for optimization throughout the study. At level two, the experimenter can apply linear programming techniques to the linear models of the different responses to determine where the best combination of responses can be found. At level three, the strategies can vary and depend largely on the nature of the fitted surfaces of the different responses and the importance of each response. For instance, if the highest priority response has a surface with a stationary region, the problem may reduce to optimizing the other responses within this region. If there are constraints to values of certain responses, the experimenter may use these to define the condition of optimization rather than an absolute extremity of the response. The point is, no hard set rule or procedure exists to define or solve the multiple response problems in the third-level of study. It may even be best to just identify a region where multiple optimum responses may exist and then to conduct experiments in this region to fit new first-order models so that linear programming techniques can be used to finish the optimization. (Linear models become better representations of response surfaces as the parameter space becomes smaller.) Nevertheless, in the end, the final set of

parameters chosen for the optimum condition depends largely on the judgement of the experimenter.

CHAPTER 5

Project Organization and Experimental Technique

In the first-level study of the reactive ion etching of polysilicon four major activities were completed:

1) Obtaining a basic understanding of the process and learning to use the equipment;

2) Developing a process and sequence of activities that would allow the polysilicon etch to be evaluated;

3) Identifying possible parameters in the process;

4) Identifying the significant parameters in the process.

The Performance Evaluation and Review Technique (PERT) chart of Figure 5.1 illustrates the precedence of these activities.[1]

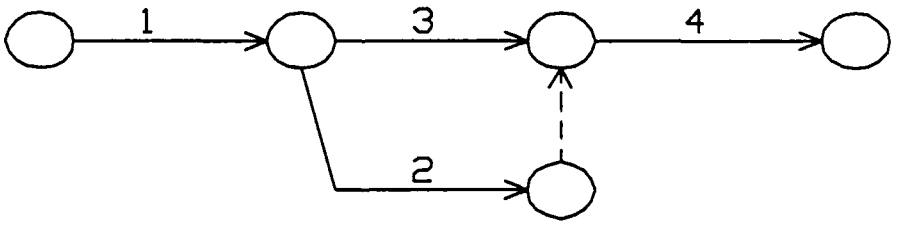


Figure 5.1 The PERT of the first-level study

1 The normal PERT chart includes activity durations and their variances. Since the PERT charts used in this chapter are only used to illustrate activity precedence, they do not include this information. The convention used in these charts is: solid arrows represent activities, dashed arrows represent dummy activities, and nodes (circles) represent events. Dummy activities are used to show precedence only and do not represent actual activities.

Chapter 1 already presented the results of activities 1, 3, and 4. Chapter 5 concentrates on the procedures established in the second activity. The chapter explains how each wafer was prepared for the etching experiments, how each wafer was processed, and how measurements were made.

WAFER PREPARATION

Wafer preparation required seven activities:

1) RCA clean wafers;
2) Oxide growth;
3) Polysilicon deposition;
4) Photolithography;
5) Polysilicon thickness measurements;
6) Photoresist thickness measurements;
7) The RIE etch of the native oxide.

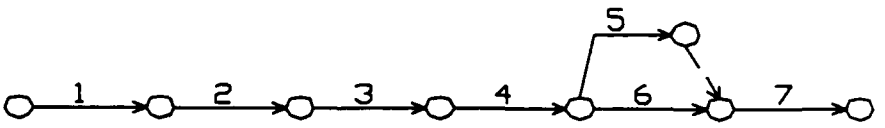


Figure 5.2 The PERT of wafer preparation

Of these seven activities, only the RCA clean was used "as is" in this study. (The clean was not modified.) All the other activities were either limited, studied, or organized to support the analysis of the etching process. The following is a summary of these activities.

OXIDE GROWTH

A nominal oxide thickness of 1000Å was selected. This oxide was grown in steam at 900° C. The actual thickness ranged from1200Å to 1280Å and from 970Å to 1020Å, depending on when processed. (The oxidation time was shortened in the second run.)

POLYSILICON GROWTH

A nominal thickness of 5000Å of undoped polysilicon was selected. The actual thickness ranged from 4200Å to 7400Å depending on the batch and the relative position of the wafer in the deposition furnace. (Different processing times were used between batches. Deposition thickness varied by as much as 1800Å from wafer to wafer in a single batch. Longer processing times were used in later runs to ensure a minimum thickness of 5000Å.)

PHOTOLITHOGRAPHY

The etching process was evaluated using the polysilicon level of the UTMOS Test mask set. This set of masks had not been used previously and, therefore, no recommended exposure times were available. A single experiment was designed to analyze the effect of exposure time on the photolithography. Reasonable results were attaianed at an exposure of 22 seconds.[2] The following summarizes the final process:

2 Later attempts to repeat these results were unsuccessful. As shown later, this significantly affected the anisotropy results of the polysilicon etch. The photolithography process needed to be optimized first.

1. A 10 minute dehydration bake at 125° C.
2. The spin application of AZ adhesion promoter at 4000 RPM for 30 seconds.
3. The spin application of Hoechst Celanese Corp. AZ 1350-SF-J photoresist at 4000 RPM for 30 seconds.
4. A 30 minute prebake at 80° C.
5. A 22 second exposure,
6. A one minute development in Shipley 452 developer.
7. The rinse in de-ionized water and dry,
8. Two 5 second duration flood exposures.
9. A 30 minute postbake at 125° C.

POLYSILICON MEASUREMENT

Polysilicon thickness measurements were made using the Nanospec/AFT optical measurement system. Measurements were made at five locations on each wafer. The exact locations were referenced to specific die and to specific locations on each die. Figure 5.3 shows their relative positions on the wafers.

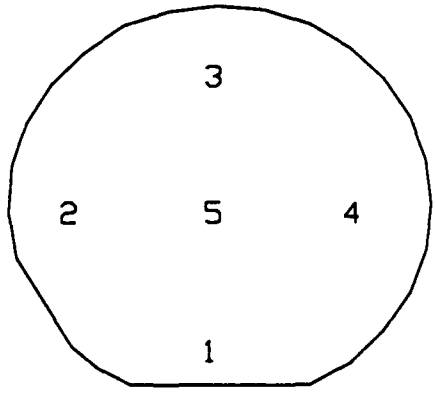


Figure 5.3 Wafer locations for film thickness measurements.

PHOTORESIST MEASUREMENT

Photoresist thickness measurements were made using a DekTak stylus measuring system. The thickness measurements were all greater than 1.6 microns. This size feature limited the accuracy of the DekTak measurement to within the nearest 200Å. This exceeded the variance in the photoresist thickness, as measured, across any single wafer. Therefore, only one measurement of the photoresist thickness was made on each wafer. [3] The measurement was made at position five.

ETCHING IN THE CF_4 PLASMA

The RIE CF_4 process was evaluated with respect to oxide, polysilicon, and photoresist etch rates. The existing recipe was left unchanged. The parameter settings and etch rates are summarized below:

SETTINGS	
CF_4 flow	50 sccm
Pressure	200 mTorr
Power (rf)	100 watts

3 Five measurements were made initially on the first 20 wafers evaluated. It was the consistent results of these measurements that led to the decision to use only one measurement.

ETCH RATES

Oxide	280 Å/min	s=32.6
Photoresist	717 Å/min	s=92.0[4]
Polysilicon	290 Å/min	s=68.0

s is the standard deviation of the sample based on the 10 measurements made at the five different locations on two wafers.

INITIAL CONDITIONS

The set of initial conditions considered in this study for the polysilicon etch were the photoresist profile and the photoresist and polysilicon thicknesses before the polysilicon etching. This set of conditions are those immediately following the RIE CF_4 process and immediately preceding the RIE $He+Cl_2$ process. Since there is a continuous transistion between these two processes, it was necessary to extrapolate the initial conditions from the measurements made during the wafer preparation and the characteristics of the CF_4 RIE.

The CF_4 etching, in this series of experiments, was used to remove the native oxide of the polysilicon. The native oxide was presumed to be less than 40Å thick. To completely etch a 40Å thick oxide layer in this process requires, on average, about 9 seconds of etching. To ensure a complete etching with a 99.99% confidence requires consideration of the process variance. Allowing a 4σ specification range, the minimum etch time becomes

4 The photoresist etch rate was evaluated using wafers with only an oxide layer beneath the photoresist. Therefore, the Nanospec was used to measure the photoresist thicknesses. These measurements were not subject to the 200 Å limit to accuracy.

13 seconds. Since it was considered extremely important that no oxide remain on the wafer, this time was further increased to 20 seconds, the time used throughout the study.

The change in the photoresist and polysilicon thickness after the CF_4 process was considered to be equivalent to the expected change after 20 seconds of RIE CF_4 etching. For photoresist, this was averaged upward to 250Å. For the polysilicon, it was averaged upward to 100Å. It should be noted that the presence of native oxide was not considered to affect the amount of polysilicon that was etched. There were three reasons for this decision: 1) The actual thickness of the native oxide was unknown; 2) The evaluation of the polysilicon etch rate was made without consideration of the native oxide; and 3) The polysilicon etch rate was essentially equal to the oxide etch rate in this process.

The photoresist profile was more difficult to determine. Measuring the photoresist profile on each wafer was impractical, since it would have involved the destruction of the wafers. (Profiles are measured using a scanning electron microscope (SEM). The wafers are cleaved across the photoresist patterns, and the cross sections are photographed in the SEM.

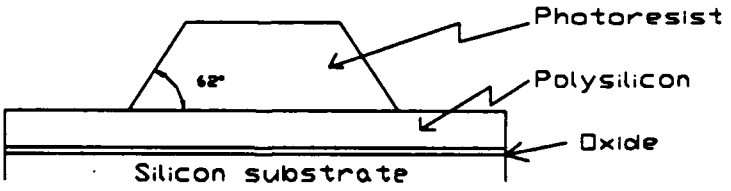


Figure 5.4 The initial photoresist profile

The profiles are then measured on these photographs.) Therefore, a major assumption was made. Since the wafers were all processed in the photolithography in the same manner, it was assumed that a single wafer could provide a representative profile of that preparatory step. Fluctuations were accounted for in the randomizing of the experiments. A test wafer was processed through the RIE CF_4 etch and was analyzed for its photoresist profile. The profile was trapezoidal and the measured angle at the photoresist polysilicon interface was 62° (See Figure 5.4).

WAFER PROCESSING AND THE MEASUREMENT OF ETCH CHARACTERISTICS

The process developed to etch all wafers was designed such that a single wafer could provide information on all etch rates, all selectivities, and the etch anisotropy. Figure 5.5 summarizes the final process in a PERT network.

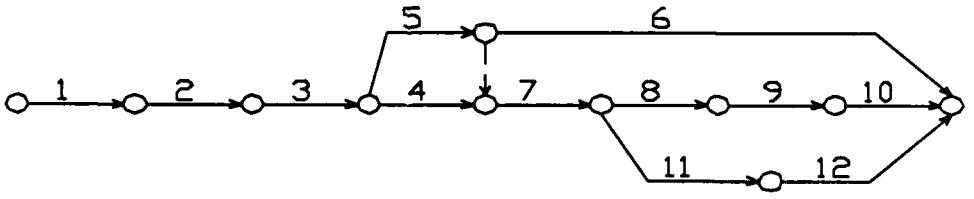


Figure 5.5 The PERT diagram of the wafer preparation

1. Partially etch the polysilicon;
2. Measure the polysilicon thickness and calculate the polysilicon etch rate;
3. Complete the Polysilicon etch;
4. Measure the oxide thickness;
5. Measure the photoresist thickness and calculate the photoresist etch rate;
6. Calculate the polysilicon/photoresist selectivity;
7. Cleave the wafer into two sections;

8. Partially etch the oxide;
9. Measure the oxide thickness and calculate the oxide etch rate;
10. Calculate the polysilicon/oxide selectivity;
11. Cleave and dissect the wafer for profile cross sections;
12. Take profile SEM photographs and measure the anisotropy.

Activity 1: Partially etch the polysilicon.

The first etch was only a partial etch of polysilicon. A process time was deliberately selected to etch only partially the polysilicon so that the etch rate could be calculated from the differences in the polysilicon thickness.

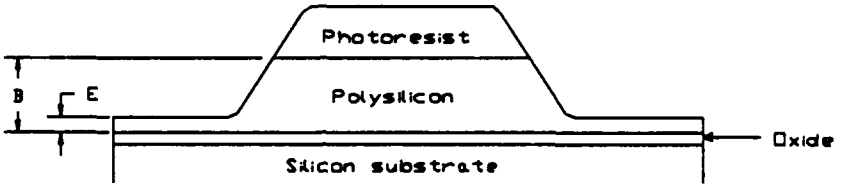


Figure 5.6 The wafer cross section after Activity 1 of the wafer processing

Activity 2: Measure the polysilicon thickness and calculate the polysilicon etch rate.

The wafer was removed from the etch chamber and the thickness of the polysilicon, E, was measured at the same locations on the wafer as those measured during the wafer preparation. The polysilicon etch rate was then calculated from these measurements and the process time, compensating for the contribution of the CF_4 process to the etch. The equation below summarizes the calculation.

$$PY_{rate} = \left(\frac{\frac{\sum_{i=1}^{5} (B_i - E_i)}{5} - 100}{t_1} \right)$$

Activity 3: Complete the polysilicon etch.

Using the largest E measured in Activity 2 and the polysilicon etch rate, a process time for etch completion was determined.. This time was increased by 5% to ensure etch completion. The equation below summarizes the calculation.

$$t_2 = \frac{E}{PYrate} \quad (1.05)$$

The wafer was then returned to the etch chamber and etched for the period t_2.

Activity 4: Measure the oxide thickness.

The oxide thickness was measured at five locations on the wafer, the same as those used in the polysilicon measurements.

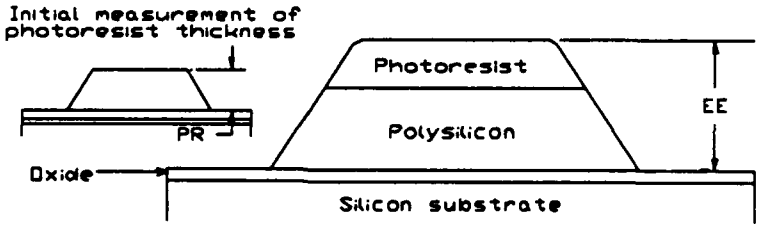


Figure 5.7 The wafer cross section after activity 3 of the wafer processing

Activity 5: Measure the photoresist thickness and calculate the photoresist etch rate.

The step height between the oxide and the top of the photoresist, EE on Figure 5.7, was measured at position 5 on the wafer. The reduction in the photoresist thickness is the difference of this measurement from the initial photoresist measurement of the wafer preparation, plus the thickness of the silicon etched.[5]

$$\Delta x_{photoresist} = (PR_5 + B_5 - EE_5),$$

The photoresist etch rate follows as

$$PR_{rate} = \frac{\Delta x_{photoresist} - 250}{(t_1 + t_2)},$$

where -250 accounts for the contribution of the initial RIE CF_4 etch.

Activity 6: Calculate the polysilicon/photoresist selectivity.

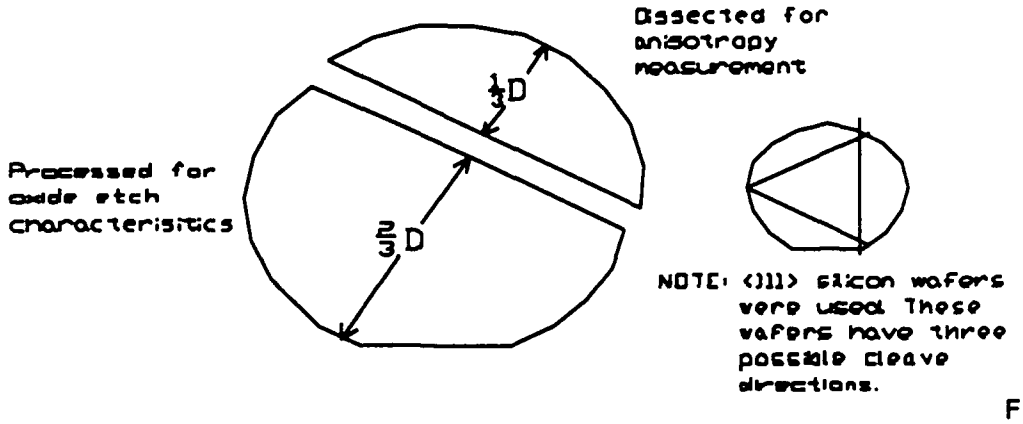
The polysilicon/photoresist selectivity, $S_{PY/PR}$ is simply the ratio of the calculated etch rates,

$$S_{PY/PR} = PY_{rate} / PR_{rate}.$$

Activity 7: Cleave the wafer into two sections.

The wafer was cleaved into two sections. One section was processed for the oxide etch evaluation and the other section was dissected for the

5 The amount of oxide etched in the second polysilicon etch due to the 5% increase in t_2 was considered very small and, therefore, was not considered. It was much less than the accuracy of the photoresist measurement, ≤ 15Å compared to 200Å.



F

Figure 5.8 The wafer cleave

anisotropy evaluation. The cleave was made along a cord at a distance 1/3 down the length of the perpendicular diameter (See Figure 5.8).

Activity 8 Partially etch the oxide.

The larger section of the wafer was then returned to the etch chamber and etched for a time, t_3. This time was deliberately selected to insure that there was only a partial etch of the oxide.

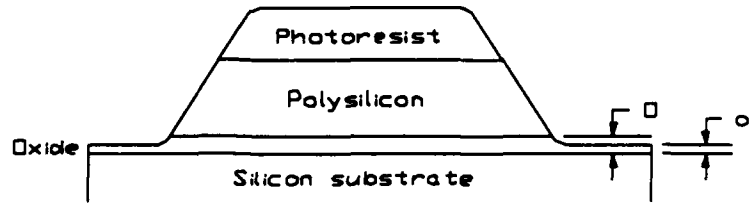


Figure 5.9 The wafer cross section after activity 8 of the wafer processing

Activity 9: Measure the oxide thickness and calculate the oxide etch rate.

The thickness of the oxide, o, was measured at the same five locations as measured in Activity 4. The oxide etch rate was calculated from these five measurements and the etch time, t_3, using the equation

$$OX_{rate} = \frac{\sum_{i=1}^{5} (O_i - o_i)}{5\, t_3}$$

Activity 10: Calculate the polysilicon/oxide selectivity.

The polysilicon/oxide selectivity, $S_{PY/OX}$, is simply the ratio of the calculated etch rates

$$S_{PY/OX} = PY_{rate} / OX_{rate} \quad .$$

Activity 11: Dissect the wafer for profile cross sections.

The smaller section of the wafer was then cleaved further into smaller slices. These slices were made such that they crossed polysilicon lines narrow enough to be viewed on both sides by the SEM. (See Figure 5.10.) Since features of this dimension were sparsely distributed across the wafer

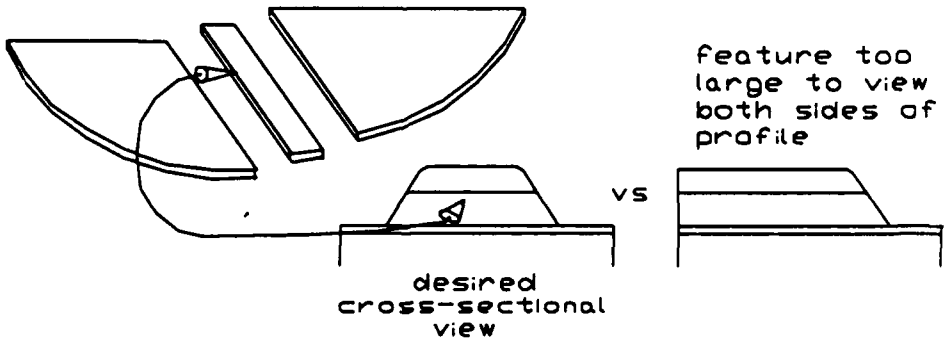


Figure 5.10 The wafer dissection and the desired cross sectional views

surface, some luck was involved in achieving a cleave that dissected one. As a result, a number of cleaves missed their mark. Therefore, the final cross sections observed under the SEM were from different locations on the wafer. But, since the intent in the dissection was to achieve a cross section on a slice near the center of the wafer (as Figure 5.10 illustrates), the distribution of cross sections was still clustered near the center. The cross sections were just not from the same wafer location.

At this point, it is helpful to digress and discuss several points of the technique. At the time of performing these experiments, the issue arose as to the best way to scribe a wafer in order to get a clean cross sectional view. In the initial attempts, the photoresist tended to peel from the polysilicon. A passing suggestion was to scribe the wafer, then freeze it in liquid nitrogen, and finally break it at the scribe. This was a dismal failure. There was enough

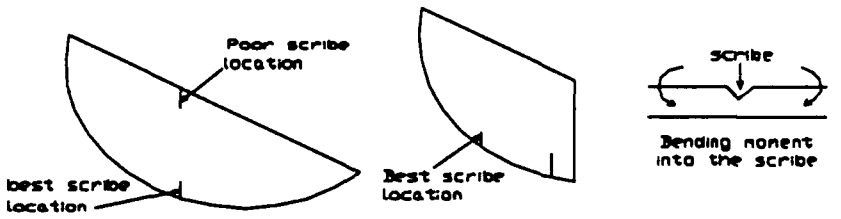


Figure 5.11 Scribe locations.

difference between the polysilicon and photoresist expansion coefficients that the photoresist literally popped-off the polysilicon and landed elsewhere on the wafer. In the end, the best way to scribe the wafer was the simplest as the following observations illustrate.

1. Scribing on the rounded and beveled edges of the wafer generates better results than scribing on the cleaved edges. (See Figure 5.11.) Scribes on the cleaved edges of the wafer tended to chip and break the wafer rather than to generate a cleave.

2. Scribes near the center of the wafer section are more successful than those near the edge of a section (See Figure 5.11).

3. The closer a scribe comes to matching the cleave direction, the better the final cleave. In many cases, the cleave results from the scribe; no bending of the wafer is necessary to obtain the cleave.

4. If it becomes necessary to bend a wafer at a scribe to cleave the wafer, the best results occur when the bending moment is away from the scribed surface (See Figure 5.11).

Activity 12: Take profile SEM photographs and measure anisotropy.

The wafer slices obtained in Activity 11 were observed directly in a SEM with no preparation (metal coating). Since the layers on these slices were insulating (photoresist and oxide), charge accumulated on the surfaces when they were exposed to the electron beam of the SEM. To minimize this effect, the SEM was used in a low acceleration voltage (1.3 keV) mode. As a

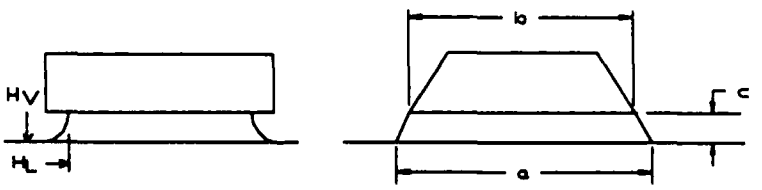


Figure 5.12 The measurements for determining etch anisotropy

result of the insulating surfaces and the low acceleration voltage, the SEM photographs were somewhat blurry but adequate for the analysis.

The anisotropy was measured and calculated directly from these cross sections. By definition, anisotropy is defined by the ratio of the lateral etch rate, H_L, measured at the polysilicon-photoresist interface and the vertical etch rate H_V, using the equation

$$A = 1 - H_L / H_V.$$

The calculation of anisotropy in these experiments required the three measurements illustrated in Figure 5.12. Each of these measurements was made five times on the photograph using a digital caliper. The measurements were repeated to account for the uncertainty in finding the true edges on the blurry photographs. Therefore, the final calculation of anisotropy was determined using the equation

$$A = \left(1 - \frac{(a-b)}{2c} \right) .$$

Photographing an accurate cross sectional view is not a trivial task. Finding and obtaining a good cross sectional view requires two translational--

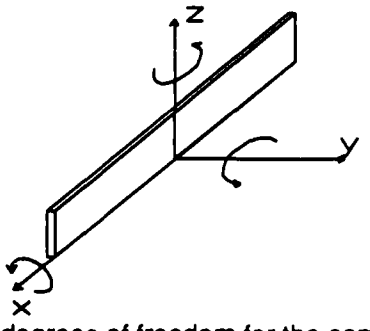


Figure 5.13 The degrees of freedom for the sample placement

x and y -- and three rotational -- x, y, and z -- degrees of freedom. The mechanism that allows the user to adjust the positioning of the sample has one too few degrees of freedom.[6] It cannot rotate about the y axis. This can be compensated for in the manner in which the samples are placed in the sample holder, but the placement is only a coarse adjustment. Since the final adjustments involve mere microns of movement, they are best done under the microscope. Failure to have a perfectly focused cross section is an indication that the plane of the cross section is not normal to the objective lens. Nevertheless, regardless of this lack of ability to rotate about the y axis, the pictures were always blurry due to the charging of the surface. Because of the blurriness, one could never be certain the cross sectional photographs were or were not accurate in the first place. It was, therefore, assumed throughout the study that the errors generated from the deficiency were random and evenly distributed.

CONCLUSION

The design of this experiment had two parts. The first part was designing a process which would allow observation of the etch characteristics (responses) that were critical to the study. An explanation of that process was the subject of this chapter. The second part of the design

6 The rotational degrees of freedom are used once the desired view is at the objective. It is at this point that the rotations are necessary. An accurate orientation of the sample is possible by displacing the sample from the axis of its rotation. Although possible, this is very impractical since it is virtually impossible to perceive the effects of these movements. Slight translational movement up and down inclines throws the sample out of focus.

was the design of a series of experiments that would allow the observation of the response surface. Chapters 2, 3, and 4 presented the theory of these designs. Chapter 6 combines these designs into a formal study of the reactive ion etch of polysilicon.

CHAPTER 6

The Response Surface Study of the Reactive Ion Etch of Polysilicon

Upon completion of the level-one study of the RIE of polysilicon, an experiment was designed to evaluate the polysilicon etching process using the current recipe as the center point of the study (center of the parameter space).[1] The chosen experimental design was a full 2^4 factorial experiment in two replicates. As will be discussed, the results of this experiment revealed that the etch recipe was in a region where the response surfaces had curvature. Therefore, the experiment was expanded to a central composite design for a subsequent third-level study. This chapter presents how these experiments were designed, conducted, and evaluated. The presentation follows the flow diagrams of the second- and third-level response surface studies presented in Chapter 4.

Designing an Experiment for a First-Order Model

A 2^4 full factorial experiment in two replicates was used in the initial experimental study of the polysilicon etching process. The current etch recipe

1 The initial intent of this experiment was to obtain a better understanding of the effects and interactions of the four parameters: Cl_2 flow rate, He flow rate, power, and pressure. It was to serve mostly as an extension of the first-level study. Fortunately, the design was adequate for the second-level study; so, this is how the design is presented in this chapter. Normally, a full 2^K factorial experiment in 2 replicates is excessively large for a second-level study.

was selected as the center of the experimental design. The parameter and controller settings for this recipe were as follows:

	PARAMETER SETTINGS	CONTROLLER SETTINGS
Chlorine flow rate	50 sccm	22.96[2]
Helium flow rate	50 sccm	14.11
Power	200 watts	0.667
Pressure	200 mTorr	200

The high and low settings of the full factorial experiment were arbitrarily selected to vary 10% from the recipe settings. Table 6.1 summarizes the high and low values of the parameters.

Table 6.1 Settings for the Full Factorial Experiment

	HIGH PARAMETER	CONTROLLER	LOW PARAMETER	CONTROLLER
Chlorine flow	55 sccm	24.750	45 sccm	19.80
Helium flow	55 sccm	16.210	45 sccm	12.67
Power	220 watts	0.733	180 watts	0.600
Pressure	220 mTorr	220.000	180 mTorr	180.0

This experimental design required 32 separate runs of the process described in Chapter 5; accordingly, 32 wafers were prepared for the experiment. These wafers were processed together in all activities of the preparatory phase except for the polysilicon deposition and the

2 The controller settings did not correspond directly to the metered readings of either gas flow rate. The chlorine setting of 22.96 corresponded to 25% of a maximum 200 sccm and the helium setting of 14.11 corresponded to 17.2% of a maximum of 290 sccm.

photolithography. The wafers were split between two runs of the polysilicon deposition, with the intent of each batch being used for one replicate of the full factorial design. The wafers were processed individually in all steps of the photolithography except in the baking activities in which they were processed in the same two batches as above. These efforts in the preparatory phase eliminated the need for blocking the experiment. But, since the photolithography was conducted a wafer at a time, it was necessary to randomize the experiment. So the experiment was conducted a replicate at a time with runs performed in a random order.[3] The wafers of the two polysilicon deposition batches were evenly and randomly distributed between the two replicates without any two runs with the same settings using wafers from the same batch. Table 6.2 summarizes this randomization effort.

Table 6.2 Experiment Order and Batch Distribution

SETTINGS (A,B,C,D)	REPLICATE 1		REPLICATE 2	
	BATCH	ORDER	BATCH	ORDER
(– 1,– 1,– 1,– 1)	A	1	B	18
(+1, – 1,– 1,– 1)	A	10	B	24
(– 1,+1, – 1,– 1)	B	2	A	30
(+1,+1, – 1,– 1)	B	6	A	26
(– 1,– 1,+1, – 1)	A	4	B	22
(+1, – 1,+1, – 1)	A	15	B	27
(– 1,+1,+1, – 1)	A	11	B	31
(+1,+1,+1, – 1)	A	5	B	28
(– 1,– 1,– 1,+1)	B	13	A	17
(+1, – 1,– 1,+1)	B	9	A	21
(– 1,+1, – 1,+1)	A	3	B	25
(+1,+1, – 1,+1)	B	7	A	20
(– 1,– 1,+1,+1)	B	16	A	23
(+1, – 1,+1,+1)	B	12	A	29
(– 1,+1,+1,+1)	B	8	A	32
(+1,+1,+1,+1)	A	14	B	19

3 Table 5 of Chapter 18 of [6:636] was used as a guide for planning this random order.

Six first-order models were obtained from the data using the method of least squares. The responses considered in these models were the polysilicon etch rate, the oxide etch rate, the photoresist etch rate, the polysilicon/oxide selectivity, the polysilicon/photoresist selectivity, and the anisotropy. Appendix 6.2 contains the data of the experiments, these models, and their analysis of variance tables. In all cases, the regression of these models exceeded a 99% significance. The significance of the lack of fit was consistently less than the significance of the regression, but it was still relatively high ranging from .90 to 0.999.

Table 6.3 The Significance of the Lack of Fit of the First-Order Models.

RESPONSE	F_{LOF}	SIGNIFICANCE	CRITERIA
Polysilicon etch rate	1.735	.90	$F_{(11,16,.90)} = 2.01$
Oxide etch rate	1.450	.90	$F_{(11,16,.90)} = 2.01$
Photoresist etch rate	2.908	.95	$F_{(11,16,.95)} = 2.46$
PY/OX Selectivity	1.225	.75	$F_{(11,48,.75)} = 1.31$
PY/PR Selectivity	4.428	.999	$F_{(11,48,.999)} = 3.92$
Anisotropy	3.287	.975	$F_{(11,16,.975)} = 2.94$

Since the lack of fit was significant for all the linear models, an additional experiment was conducted at the experiment's center to verify that curvature did, in fact, exist. Table 6.4 shows a comparison of the measured results and the average of the full factorial experiment.

Table 6.4 Comparison of the Center results and the Linear Model Prediction

RESPONSE	EXPERIMENT CENTER	MODEL CENTER
Polysilicon etch rate	2877	2838 Å/min
Oxide etch rate	83.7	125 Å/min
Photoresist etch rate	2655	2763 Å/min
PY/OXSelectivity	34.37	24.04
PY/PRSelectivity	1.084	1.043
Anisotropy	0.360[4]	0.5422

Some of the measured responses were very close to the responses predicted by their respective models, but the two most important responses, Polysilicon/Oxide selectivity and anisotropy, were very different.[5] The high significance of the lack of fit of the linear models and the evidence of curvature in the two most critical responses of the process warranted an increase in the level of the response surface study.

4 The wafer used for this center experiment came from a different batch than those used in the full factorial design. Characteristics of this batch, especially of the photoresist profile, were much different. This drastically affected the anisotropy. So, the measurements of anisotropy on the wafers of this batch were adjusted to account for the different initial conditions. The value shown in this chart, 0.360, is the adjusted value. The unadjusted value was 0.0051. Appendix 6.4 explains how this value was adjusted.

5 A better way to make this analysis would include several runs of the center experiment. The standard deviation of the responses of these runs would allow a statistical analysis of whether the measured and the predicted responses are the same using the t distribution statistic.

EXTENDINGING THE EXPERIMENT TO A SECOND-ORDER DESIGN

The full factorial experiment was extended to a central composite design. The center and axial point experiments were each conducted once. The second-order models of the polysilicon etch rate, the oxide etch rate, the photoresist etch rate, the polysilicon/oxide selectivity, the polysilicon/photoresist selectivity, and the anisotropy[6] were fitted to the data. Appendices 6.3a-f have the first fitted models to this data. These appendices include the model coefficients, the analysis of variance for the models, the canonical analysis information, and the canonical analysis table.

ANALYZING THE FIT OF THE SECOND-ORDER MODELS

The analysis of variance tables of all the new models revealed that the regression was significant to better than 99.9% in all cases. It was not significantly different from that of the linear models, however. The lack of fit F statistics decreased in all cases except the polysilicon/oxide selectivity ($S_{PY/OX}$) and the anisotropy (A) models. This increase in the lack of fit is rather ironic considering that the curvature of the latter two responses in the full factorial experiment was the motivation for changing to a larger experiment. But, nevertheless, the general trend toward improvement is adequate evidence that the second-order models better represent the data. The lack of improvement in the $S_{PY/OX}$ and anisotropy models can be attributed to other factors, which will be discussed later.

6 The anisotropy measurements of the center and axial experiments were affected by the initial photoresist profile on these wafers. The data used to fit this model was adjusted to account for this discrepancy. Appendix 6.4 explains this adjustment.

Table 6.5 The Comparison of F the Statistics for the First- and Second-Order Models

RESPONSE	FIRST-ORDER (R|ß 0)	FIRST-ORDER LOF	SECOND-ORDER (R| ß0)	SECOND-ORDER LOF
PYETCHRATE	6.060e+01 [1]	1.735[2]	2.091e+01 [5]	1.149[6]
0XETCHRATE	1.280e+02 [1]	1.450[2]	4.635e+01 [5]	1.372[6]
PRETCHRATE	1.163e+02 [1]	2.908[2]	4.028e+01 [5]	1.538[6]
SPY/OX	6.299e+01 [3]	1.225[4]	2.397e+01 [7]	2.511[8]
SPY/PR	2.594e+01 [3]	4.428[4]	1.056e+01 [7]	1.898[8]
A	1.329e+01 [1]	3.287[2]	1.517e+01 [5]	7.385[6]

The superscripts on the F statistics above refer to the rows below. Each row lists the degrees of freedom of the mean squares used to form the specific F statistics and includes a table of F statistics for the evaluation of significance.

	DOF_1	DOF_2	.75	.90	.95	.975	.99	.995	.999
1	4	16	1.50	2.33	3.01	3.73	4.77	5.64	7.94
2	11	16	1.44	2.01	2.46	2.94	3.62	4.19	5.68
3	4	48	1.38	2.07	2.57	3.08	3.91	4.26	5.55
4	11	48	1.31	1.78	2.00	2.28	2.65	2.93	3.92
5	14	16	1.42	1.96	2.37	2.82	3.46	3.98	5.36
6	10	16	1.44	2.03	2.49	2.99	3.69	4.27	5.81
7	14	48	1.29	1.65	1.90	2.16	2.50	2.75	3.33
8	10	48	1.32	1.73	2.04	2.34	2.72	3.03	3.74

Since there is potential for improvement in these models the next step is to attempt variable recoding.

IMPROVING THE MODEL'S FIT

It was apparent from the large F statistics for the lack of fit that the models could be improved. To improve the models' fit, several techniques of

encoding the variables and responses were attempted. This section attempts to explain how the variables were encoded differently and the motivation for each change.

THE BASE ENCODING TECHNIQUE.

The experiments were both designed and first analyzed using a linear encoding procedure. This procedure is explained on page 23 of Chapter 2. It simply involves using the high and low values of the full factorial design to specify a line and then using this line and the desired encoded values to determine the parameter settings. It is summarized in the equation

$$x_i = 2\ (p_i - \bar{p}_i) / d_i$$

The response measurements, as explained in Chapter 5, were simply those most normally associated with the defineition of the responses.

1. Using an Angular Measurement of Anisotropy.

Anisotropy, as normally defined, is

$$1 - H_L/H_V \qquad (*)$$

where H_L and H_V are the lateral and vertical etch rates, respectively. As Figure 6.1 illustrates, the cross sections of the etch profiles indicate that the angle of the polysilicon profile can provide an equivalent anisotropy measure. Note that the equation (*) above is equivalent to

$$1 - \cot \theta.$$

Therefore, this response can be reencoded as a pure θ where 90° is an optimum, or as a different trigonometric function of θ with the optimum depending on the function chosen.

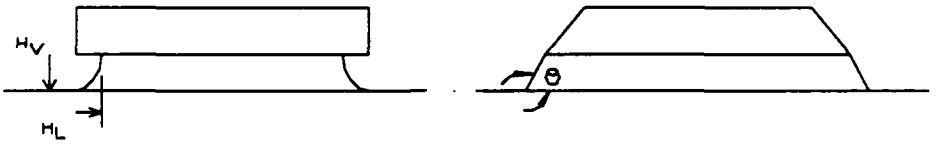


Figure 6.1 The Angular Evaluation of Anisotropy.

2. Encoding Power.

The basic definition of power is flow times effort, which in electric circuits is current times voltage. In the plasma, current is achieved in the transport of charged particles. The transport of these particles helps define the availability of the reactant species at the wafer surface. So, the nature of the etch is in some way proportional to the current I. Similarly, in a plasma, the voltage potential across the electrodes defines the energy of the charged species. The final energy of the species determines its effectiveness in either the sputtering or the damaging of the wafer surface. So, here the nature of the etch is in some way proportional to the voltage. Since power is related to the square of current and the square of voltage, it is reasonable to assume the existence of mechanisms that are functions of the square root of power. Since there is obvious coupling of the flow and voltage, the power term cannot be neglected. So, in a general method of reencoding the power parameter, a weighting function that linearly combined the power and the square root of power was used:

$$C = (w_1(P)^{1/2} + w_2P).$$

In this function, $(w_1 + w_2)$ is considered equal to 1. The equation for encoding the parameters then becomes

$$x_3 = 2\,(C - \overline{C})\,/\,d$$

where

$$\overline{C} = \frac{(w_1(220)^{1/2} + w_2(220) + (w_1(180)^{1/2} + w_2180)}{2}$$

and

$$d = \frac{(w_1(220)^{1/2} + w_2220) - (w_1(180)^{1/2} + w_2180)}{2} .$$

Table 6.6 The Encoded Power Parameters

NUM	w_1	w_2	$AXIAL_H$	HIGH	CENTER	LOW	$AXIAL_L$
1	0.00	1.00	1.4140	1.0000	0.0000	-1.0000	-1.4140
2	0.10	0.90	1.4139	1.0000	0.0001	-1.0000	-1.4141
3	0.20	0.80	1.4138	1.0000	0.0002	-1.0000	-1.4142
4	0.30	0.70	1.4137	1.0000	0.0004	-1.0000	-1.4144
5	0.40	0.60	1.4135	1.0000	0.0006	-1.0000	-1.4146
6	0.50	0.50	1.4132	1.0000	0.0009	-1.0000	-1.4149
7	0.60	0.40	1.4128	1.0000	0.0013	-1.0000	-1.4154
8	0.70	0.30	1.4122	1.0000	0.0019	-1.0000	-1.4161
9	0.80	0.20	1.4111	1.0000	0.0031	-1.0000	-1.4173
10	0.90	0.10	1.4083	1.0000	0.0061	-1.0000	-1.4205
11	1.00	0.00	1.3906	1.0000	0.0250	-1.0000	-1.4410

3. Encoding Pressure.

The pressure, as Chapter 1 describes, can affect the etch process by increasing the absorption rate of the chlorine at the polysilicon surface as well as affect the transport of species in the plasma. Pressure directly affects the composition of the plasma and the rate at which reactions occur. The

physical factors that pressure affects which most directly influence the plasma composition and the chlorine absorption are the gas density in the chamber and the mean free path of the gas molecules. Of course, the gas density is directly proportional to the pressure. The mean free path affects the plasma in two ways: It affects the residence time of the reactant species at the wafer surface; and it affects the rate at which the reactant species move to the wafer surface. The mean free path is inversely proportional to the pressure. Since residence time is inversely proportional to the mean free path, it follows that the residence time is directly proportional to the pressure. The rate at which the reactant species moves to the wafer surface is dependent on the probability that it will undergo a collision in the distance it must travel. In the case of a gas with a mean free path of λ, the probability that it will have a collision in a distance d is

$$n/n_0 = \exp(-d/\lambda).$$

Pressure, therefore, also has an exponential relationship to the etch,

$$n/n_0 = (K)^p$$

where p is pressure and K is the constant when d and all the other variables that affect λ are held constant. So, in a manner similar to that used in encoding power the pressure was encoded using the equation,

$$D = (w_1\ 20\ \exp(p/100) + w_2\ p).$$

The constants of 20 and 100 were used in this equation to allow the weighting terms, w_1 and w_2, to determine which pressure terms were dominant. A second method attempted in encoding the pressure used a combination of the natural log of pressure and pressure with the equation

$$D = w_1\ \ln(p) + w_2\ (p/50).$$

Table 6.7 The Encoded Pressure Parameters

FIRST METHOD

NUM	w_1	w_2	$AXIAL_H$	HIGH	CENTER	LOW	$AXIAL_L$
1	0.00	1.00	1.4140	1.0000	0.0000	-1.0000	-1.4140
2	0.10	0.90	1.4296	1.0000	-0.0141	-1.0000	-1.4011
3	0.20	0.80	1.4437	1.0000	-0.0270	-1.0000	-1.3894
4	0.30	0.70	1.4567	1.0000	-0.0388	-1.0000	-1.3786
5	0.40	0.60	1.4686	1.0000	-0.0496	-1.0000	-1.3688
6	0.50	0.50	1.4796	1.0000	-0.0596	-1.0000	-1.3597
7	0.60	0.40	1.4897	1.0000	-0.0688	-1.0000	-1.3513
8	0.70	0.30	1.4992	1.0000	-0.0774	-1.0000	-1.3435
9	0.80	0.20	1.5079	1.0000	-0.0853	-1.0000	-1.3362
10	0.90	0.10	1.5161	1.0000	-0.0927	-1.0000	-1.3295
11	1.00	0.00	1.5237	1.0000	-0.0997	-1.0000	-1.3231

SECOND METHOD

NUM	w_1	w_2	$AXIAL_H$	HIGH	CENTER	LOW	$AXIAL_L$
12	0.00	1.00	1.4140	1.0000	0.0000	-1.0000	-1.4140
13	0.10	0.90	1.4116	1.0000	0.0026	-1.0000	-1.4169
14	0.20	0.80	1.4089	1.0000	0.0056	-1.0000	-1.4202
15	0.30	0.70	1.4059	1.0000	0.0089	-1.0000	-1.4238
16	0.40	0.60	1.4025	1.0000	0.0126	-1.0000	-1.4279
17	0.50	0.50	1.3987	1.0000	0.0167	-1.0000	-1.4325
18	0.60	0.40	1.3943	1.0000	0.0215	-1.0000	-1.4378
19	0.70	0.30	1.3893	1.0000	0.0270	-1.0000	-1.4438
20	0.80	0.20	1.3834	1.0000	0.0334	-1.0000	-1.4509
21	0.90	0.10	1.3765	1.0000	0.0410	-1.0000	-1.4593
22	1.00	0.00	1.3682	1.0000	0.0501	-1.0000	-1.4693

4. Encoding Flow Rates

The flow rates of the gases affect the process in two ways. First of all, it is through the flows that the gases are replenished, but the flows also affect how well the by-products of the etching reactions are removed from the wafer surface. Characterization of this latter mechanism is not easy, since it is independent of the type of gas or how the gas is consumed in the process.

The effect of moving etch by-products with gas flows is a linear combination of the flow rates of both the chlorine and helium gases. Additionally, the nature of the mechanism is not readily recognizable either. If flow is laminar across the wafer surface, one might suspect that a stagnant layer is formed through which reactants must diffuse. If the flow is turbulent, this layer would not occur in the same way. So, the trial encoding methods of the gas flows were true trial and error efforts. The attempted encoding equations were based on applying the equations

$$A \text{ or } B = (w_1 \ln(\text{flow}) + w_2 (\text{flow})),$$

$$A \text{ or } B = (w_1 \text{ sqrt}(\text{flow}) + w_2 (\text{flow})),$$

and

$$A \text{ or } B = (w_1 \, 6 \exp(\text{flow}/25) + w_2 (\text{flow})).$$

Table 6.8 The Encoded Flow Rates

FIRST METHOD

NUM	w_1	w_2	$AXIAL_H$	HIGH	CENTER	LOW	$AXIAL_L$
1	0.00	1.00	1.4140	1.0000	0.0000	-1.0000	-1.4140
2	0.10	0.90	1.4139	1.0000	0.0001	-1.0000	-1.4141
3	0.20	0.80	1.4138	1.0000	0.0003	-1.0000	-1.4143
4	0.30	0.70	1.4136	1.0000	0.0004	-1.0000	-1.4145
5	0.40	0.60	1.4134	1.0000	0.0007	-1.0000	-1.4147
6	0.50	0.50	1.4131	1.0000	0.0010	-1.0000	-1.4151
7	0.60	0.40	1.4127	1.0000	0.0015	-1.0000	-1.4156
8	0.70	0.30	1.4120	1.0000	0.0022	-1.0000	-1.4165
9	0.80	0.20	1.4106	1.0000	0.0037	-1.0000	-1.4181
10	0.90	0.10	1.4070	1.0000	0.0077	-1.0000	-1.4225
11	1.00	0.00	1.3682	1.0000	0.0501	-1.0000	-1.4693

Table 6.8 Gas Flow Rates (continued)

SECOND METHOD

NUM	w_1	w_2	$AXIAL_H$	HIGH	CENTER	LOW	$AXIAL_L$
12	0.00	1.00	1.4140	1.0000	0.0000	-1.0000	-1.4140
13	0.10	0.90	1.4138	1.0000	0.0002	-1.0000	-1.4142
14	0.20	0.80	1.4136	1.0000	0.0004	-1.0000	-1.4145
15	0.30	0.70	1.4133	1.0000	0.0007	-1.0000	-1.4148
16	0.40	0.60	1.4129	1.0000	0.0011	-1.0000	-1.4152
17	0.50	0.50	1.4125	1.0000	0.0017	-1.0000	-1.4158
18	0.60	0.40	1.4118	1.0000	0.0024	-1.0000	-1.4166
19	0.70	0.30	1.4107	1.0000	0.0036	-1.0000	-1.4178
20	0.80	0.20	1.4088	1.0000	0.0055	-1.0000	-1.4200
21	0.90	0.10	1.4049	1.0000	0.0097	-1.0000	-1.4245
22	1.00	0.00	1.3906	1.0000	0.0250	-1.0000	-1.4410

THIRD METHOD

NUM	w_1	w_2	$AXIAL_H$	HIGH	CENTER	LOW	$AXIAL_L$
23	0.00	1.00	1.4140	1.0000	0.0000	-1.0000	-1.4140
24	0.10	0.90	1.4322	1.0000	-0.0165	-1.0000	-1.3990
25	0.20	0.80	1.4478	1.0000	-0.0308	-1.0000	-1.3860
26	0.30	0.70	1.4615	1.0000	-0.0432	-1.0000	-1.3746
27	0.40	0.60	1.4736	1.0000	-0.0542	-1.0000	-1.3646
28	0.50	0.50	1.4843	1.0000	-0.0639	-1.0000	-1.3558
29	0.60	0.40	1.4939	1.0000	-0.0726	-1.0000	-1.3478
30	0.70	0.30	1.5025	1.0000	-0.0804	-1.0000	-1.3407
31	0.80	0.20	1.5102	1.0000	-0.0874	-1.0000	-1.3343
32	0.90	0.10	1.5173	1.0000	-0.0938	-1.0000	-1.3285
33	1.00	0.00	1.5237	1.0000	-0.0997	-1.0000	-1.3231

The different encoding methods were evaluated by comparing the F statistics for the lack of fit F_{LOF} for the three models $S_{PY/OX}$, $S_{PY/PR}$ and Anisotropy. In performing these evaluations, only the parameter in question was reencoded; all other parameters remained at their original encoded values. In each of the different methods of reencoding the parameters, it was

observed that F_{LOF} changed continuously, either increasing or decreasing as the skew of the reencoded variables moved to one side or the other.[7] In the case of the chlorine, power, and pressure variables the direction of the change of F_{LOF} for the three models evaluated was the same. The greatest improvements came when chlorine was reencoded as 33 of Table 6.8, when power was encoded as 11 of Table 6.6, and when pressure was reencoded as 22 of Table 6.7. In the case of the helium flow variable, the F_{LOF} changed in different ways for the models. It would either increase for $S_{PY/OX}$ as it decreased for $S_{PY/PR}$ and anisotropy, or vice versa. Therefore, the original experimental method of encoding was retained for the helium variable.

To determine the best combination of reencoded variables, all possible combinations of these variables and the original encoded variables were evaluated for their effect on the F_{LOF} for all of the models. This included the response of anisotropy when it was measured by its profile angle. Table 6.9 summarizes these results.

7 Skew in this context refers to how the axial variables change. A skew exists if |Axial High| ≠ | Axial Low| . A set of encoded variables with |Axial High|<|Axial Low| has the opposite skew of a set of variables with |Axial High| >|Axial Low|.

Table 6.9 F_{LOF} Statistics for All Responses with Reencoded Variables[8]

PARAMETER ENCODED	POLY ETCH	OXIDE ETCH	PR ETCH	$S_{PY/OX}$	$S_{PY/PR}$	A	A_q[9]
	1.149	1.372	1.538	2.511	1.898	7.385	5.225
A	1.223	1.444	1.601	2.496	1,895	7.570	5.349
C	1.151	1.381	1.536	2.500	1.892	7.301	5.168
AC	1.225	1.451	1.606	2.486	1.880	7.484	5.274
D	1.149	1.305	1.555	2.444	1.890	7.196	5.090
AD	1.221	1.370	1.616	2.425	1.885	7.379	5.214
CD	1.151	1.314	1.553	2.432	1.882	7.112	5.034
ACD	1.223	1.378	1.616	2.414	1.870	7.293	5.137

As can be seen, the improvements in F_{LOF} for the different models were not the same. In some cases the models improved while in others they decreased. Generally speaking, the effect of reencoding power and pressure was an improved model. Reencoding the chlorine flow had mixed results. Due to its consistent effect of decreasing the fit of the etch rate models, it was considered a poor encoding technique. Therefore, only the power and pressure variables were reencoded for the models used in the analysis. Additionally, the model of anisotropy, with the response encoded as an angle was used in the analysis due to its nearly 30% better F_{LOF}. Appendices 6.3g-l include the complete analysis of these models, including the canonical analysis data.

8 *This table compares the F statistics for all possible combinations of* recoding using two possibilities for the parameters A, C, and D. The different combinations are identified by the column labeled "Parameter Encoded." For example, the row identified by AD lists the F statistics of all the models when the parameters A and D, the chlorine flow rate and the pressure, were reencoded, and the codings of parameters B and C, the helium flow rate and the power, were left unchanged.

9 Anisotropy is encoded as an angle.

DEVELOPING A STRATEGY TO FIND THE OPTIMUM CONDITION

In a situation where multiple responses must be simultaneously optimized, the experimenter must simultaneously consider the importance of each response and the adequacy of each model. In this experiment, if only the importance of each response were considered, then the anisotropy would have been the response optimized. But, since the adequacy of the anisotropy model was questionable, this strategy was reconsidered.

The F_{LOF} increased for the two important responses of anisotropy and $S_{PY/OX}$ when the first-order models were expanded to second-order models. In the case of the anisotropy model, the increase in F_{LOF} can be attributed to biasing elements encountered when the full factorial experimental design was expanded to a central composite experimental design. In other words, the model was grossly influenced by a fifth parameter, the initial photoresist profile, which changed for the expanded portion of the experiment. The anisotropy model was not very trustworthy because of this influence. In the case of the $S_{PY/OX}$ model, however, the increase in F_{LOF} did not seem to be the result of biasing elements. This conclusion is supported by the low F_{LOF} statistic for both the oxide and polysilicon etch rate models. The increase in the F_{LOF} for the $S_{PY/OX}$ model must, therefore, be attributed to the existence of a higher order type of curvature. This conclusion can be justified in observing the response surfaces of the oxide and polysilicon etch rate models. The oxide etch rate model is a valley and the polysilicon etch rate model is a saddle. Additionally, these models have stationary points that are far apart. The significance of these two types of surfaces is that the surface

formed by their ratio is very likely to have multiple stationary points, a condition that results in a higher order surface. So, the high F_{LOF} statistic for the $S_{PY/OX}$ model does not discredit the use of the model in predicting where optimum conditions may be found, as is the case with the anisotropy model.

The strategy used to optimize the process concentrated on analyzing the $S_{PY/OX}$ and the $S_{PY/PR}$ models. The $S_{PY/OX}$ model was selected for obvious reasons; it is the model of one of the responses that I desired to optimize. The second model, $S_{PY/PR}$ was selected as an alternative to the anisotropy model since the $S_{PY/PR}$ response combined with the initial profile of the photoresist seemed to be the dominant factors contributing to the final anisotropy. An improved $S_{PY/PR}$ response almost always led to an improved etch anisotropy.

The $S_{PY/OX}$ model was specifically well suited for an optimizing strategy. As the data of Appendix 6.3j shows, this model is a saddle. It has one direction of relative maxima and three directions of relative minima. But, this is a global analysis. In considering the region in the vicinity of the experimental design, only one feature is present: the relative maxima. (This is readily identified in the canonical analysis table. The feature with positive curvature is only a distance of 0.661 from the center of the experimental design, while all of the other features are more than a distance of 4.4 from the center.) Since this relative maxima is the sole relative maxima, it can be considered a ridge. Additionally, since the magnitude of the curvature is small, the ridge can be considered stationary. This provides a region over which one may choose parameter settings to achieve multiple optimum

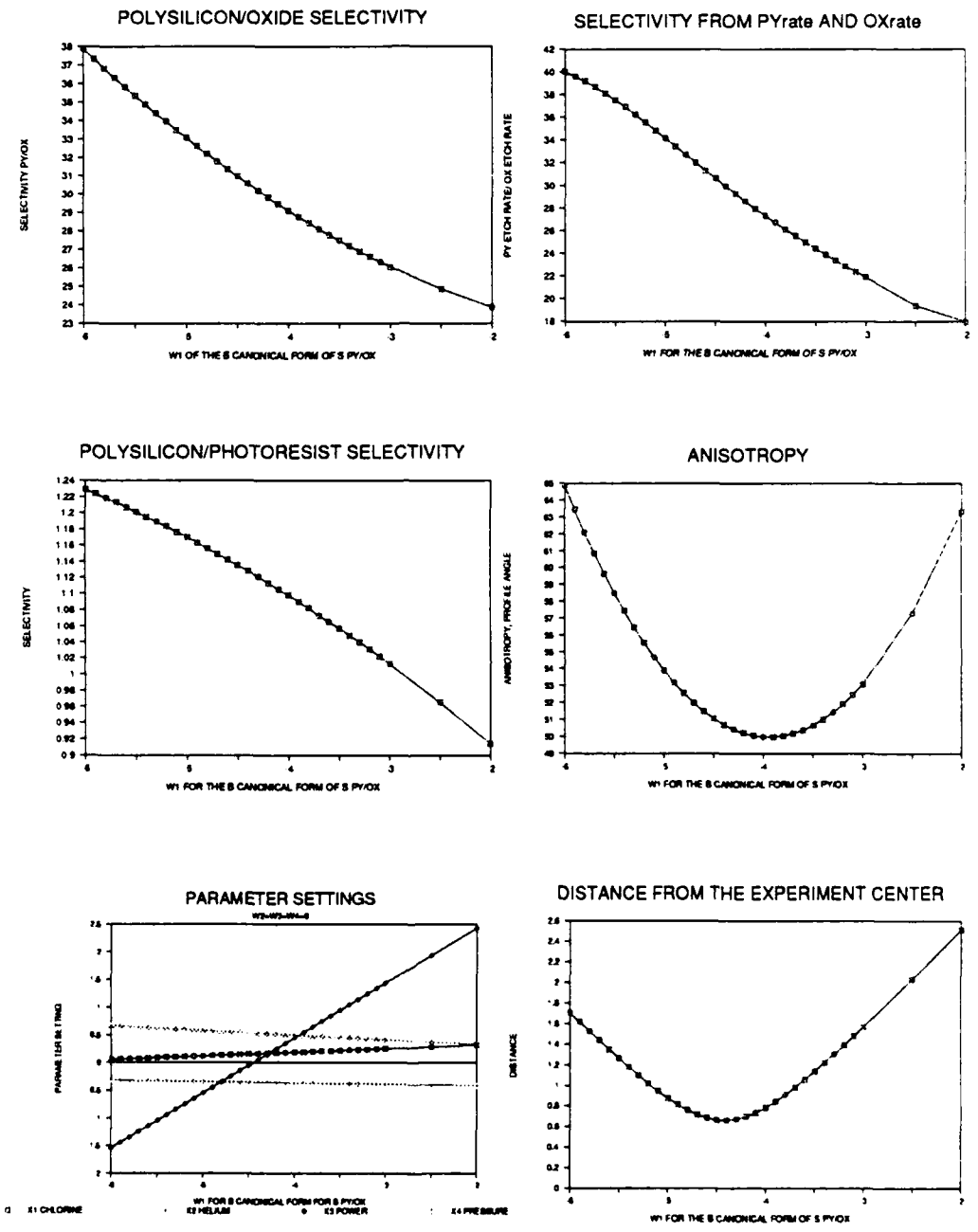


Figure 6.2 The Analysis of the Ridge Feature of the $S_{PY/OX}$ Model.

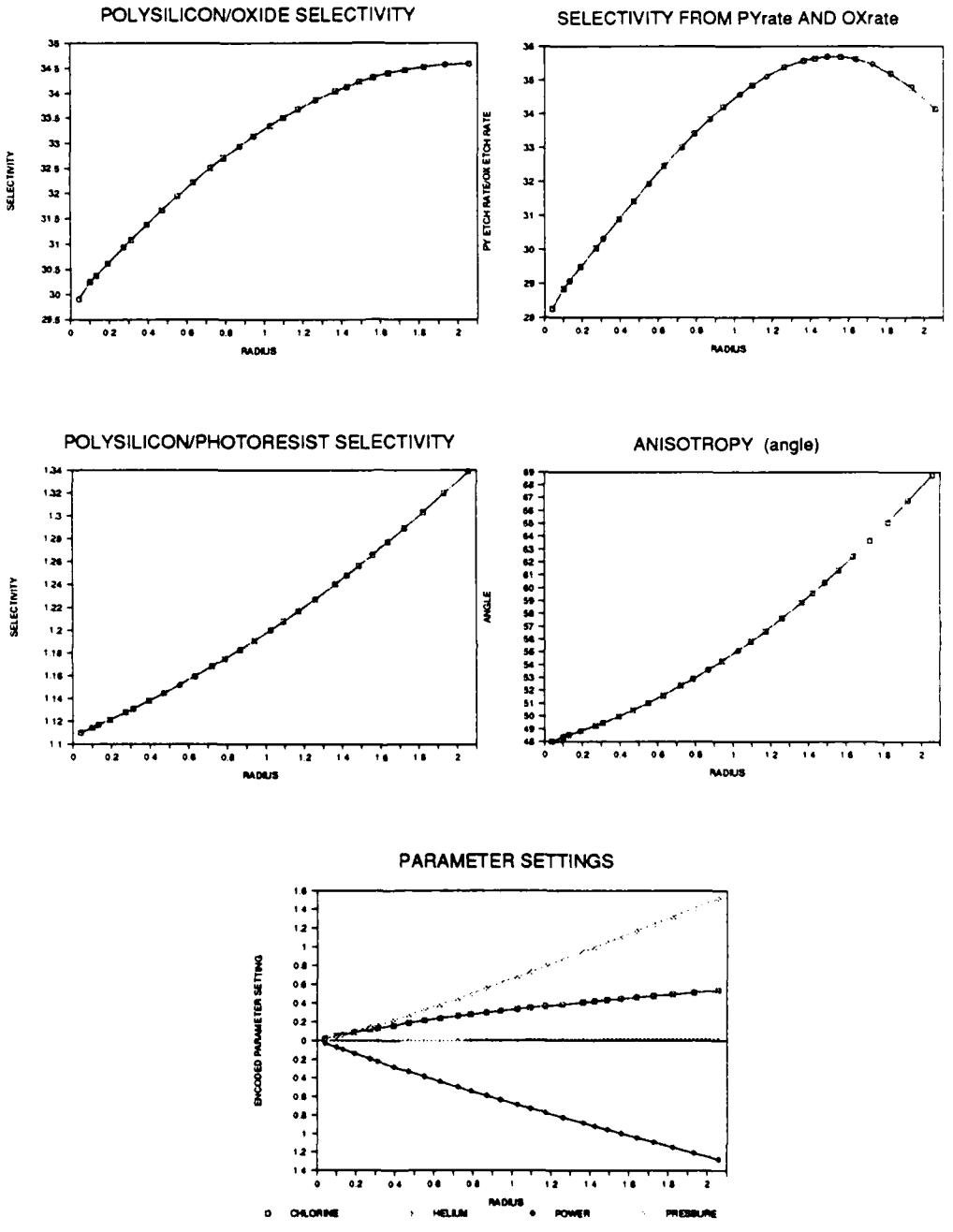


Figure 6.3 The Ridge Analysis of the SPY/PR Model.

responses. Figure 6.2 shows graphs of all the responses to positions along this ridge.

The $S_{PY/PR}$ model was not well suited for optimization. It too was a saddle, but there was no clear optimum feature. Therefore, the method of ridge analysis was selected as the method to determine the directions of relative maxima in the experiment. Figure 6.3 illustrates the changes of the different responses along the path identified in the ridge analysis of the $S_{PY/PR}$ model.

There were common trends in the results of these two procedures. As the responses improved the power decreased, the helium flow rate remained somewhat constant, the chlorine flow rate changed only a small amount, and the pressure generally increased. Additionally, the important responses, anisotropy, $S_{PY/OX}$, and $S_{PY/PR}$, all increased together. Finally, the two procedures provided paths of relative maxima that both progressed to the same region of the parameter space. Table 6.10 lists the points, at a distance of 1.5 from the experiment's center, for the two procedures. These points are

Table 6.10 The Points of Maximum Response at the Distance 1.5.

		$S_{PY/OX}$	$S_{PY/PR}$
Chlorine Flow	x_1	0.066	0.433
Helium Flow	x_2	-0.326	0.013
Power	x_3	-1.342	-0.960
Pressure	x_4	0.645	1.051

only a distance of about 0.75 from each other . These trends are promising and indicate the existence of a good region to explore for an optimum condition for all the responses.

At this point, I concluded my study of the reactive ion etch of polysilicon. As is evident, additional experimentation is needed to be certain of the location of the optimum responses. Therefore, I present recommendations on additional experiments to conduct. It is important to note that I deviated from the flow charts of the level-three study presented in Chapter 4. The flow chart of Figure 4.13 is for the study of a single response. If I were only optimizing $S_{PY/OX}$, I would conduct experiments along the maxima identified by the ridge and verify that the optimum condition existed. If I were optimizing $S_{PY/PR}$, I would conduct experiments along the path identified by the method of ridge analysis until improvement stopped, and would then design a new experiment about that point. But in this case since both models identify the same region for the optimum response, I believe it is best to begin with the design of a new experiment about that region. The region that should be evaluated in encoded variables is

$$0.000 \leq x_1 \leq 0.500$$
$$-0.400 \leq x_2 \leq 0.100$$
$$-1.500 \leq x_3 \leq -0.750$$
$$0.500 \leq x_4 \leq 1.250,$$

which is equivalent to

$$50.00 \text{ sccm} \leq a \leq 52.50 \text{ sccm}$$

$$48.00 \text{ sccm} \leq b \leq 50.50 \text{ sccm}$$

$$170.6 \text{ watts} \leq c \leq 184.8 \text{ watts}$$

$$214.5 \text{ mTorr} \leq d \leq 225.6 \text{ mTorr.}$$

These settings should become the low and high settings of a new experimental design. Since this experimental region would be spatially smaller than the one evaluated in the first design, one could not be certain that curvature was a factor. For this reason, I recommend returning to the beginning of the level-two study flow chart of Figure 4.12.[10] Additionally, I recommend beginning the study with a resolution IV fractional factorial design.

10 Since the parameter space is smaller, a linear model may be adequate in modeling the response.

APPENDIX 6.1a. The Parameter Significance to the Polysilicon Etch Rate

The parameters were evaluated in a full factorial experimental design.

PARAMETER	DESCRIPTION	HIGH VALUE	LOW VALUE	UNIT OF MEASURE
A	Chlorine flow	55	45	sccm
B	Helium flow	55	45	sccm
C	Power	220	180	watts
D	Pressure	220	180	mTorr

The response unit of measure was Å polysilicon/minute.

ENCODED PARAMETERS (A,B,C,D)	RESPONSE OF REPLICATE 1	2	CONTRASTS
(– 1,– 1,– 1,– 1)	2228.00	2600.00	90812.98
(+1, – 1,– 1,– 1)	2585.00	2571.50	– 203.88
(– 1,+1, – 1,– 1)	2565.00	2693.45	1401.62
(+1,+1, – 1,– 1)	2430.00	2665.00	908.32
(– 1,– 1,+1, – 1)	3269.00	3320.00	12188.72
(+1, – 1,+1, – 1)	3174.00	3445.00	– 117.98
(– 1,+1,+1, – 1)	3291.00	2837.40	– 1168.28
(+1,+1,+1, – 1)	3293.40	3313.27	1304.22
(– 1,– 1,– 1,+1)	2260.00	2380.00	– 1749.42
(+1, – 1,– 1,+1)	2158.00	2231.00	– 1550.52
(– 1,+1, – 1,+1)	2474.00	2477.00	1609.58
(+1,+1, – 1,+1)	2406.00	2588.00	995.68
(– 1,– 1,+1,+1)	3249.00	3324.80	978.48
(+1, – 1,+1,+1)	2852.00	3058.20	– 804.42
(– 1,+1,+1,+1)	3370.00	3169.60	515.48
(+1,+1,+1,+1)	3267.00	3267.00	– 576.22

EFFECT	SUM OF SQUARES	DOF	MEAN SQUARES	F_{STAT}	SIGNIFICANCE	ESTIMATE OF EFFECT
1	0.00	1	0.00	0.0000	>99.9	5675.79
A	1298.97	1	1298.97	0.0656	<75.0	- 12.74
B	61391.83	1	61391.83	3.0998	>90.0	87.60
AB	25782.66	1	25782.66	1.3018	<75.0	56.77
C	4642652.98	1	4642652.98	234.4158	>99.9	761.79
AC	434.98	1	434.98	0.0220	<75.0	- 7.37
BC	42652.44	1	42652.44	2.1536	>75.0	- 73.02
ABC	53155.93	1	53155.93	2.6839	>75.0	81.51
D	95639.70	1	95639.70	4.8290	>95.0	- 109.34
AD	75128.51	1	75128.51	3.7934	>90.0	- 96.91
BD	80960.87	1	80960.87	4.0879	>90.0	100.60
ABD	30980.58	1	30980.58	1.5643	>75.0	62.23
CD	29919.47	1	29919.47	1.5107	>75.0	61.15
ACD	20221.61	1	20221.61	1.0210	<75.0	- 50.28
BCD	8303.74	1	8303.74	0.4193	<75.0	32.22
ABCD	10375.92	1	10375.92	0.5239	<75.0	- 36.01
ERROR	316883.33	16	19805.21			
TOTAL	5495783.53	31				

APPENDIX 6.1b. The Parameter Significance to the Oxide Etch Rate

The parameters were evaluated in a full factorial experimental design.

PARAMETER	DESCRIPTION	HIGH VALUE	LOW VALUE	UNIT OF MEASURE
A	Chlorine flow	55	45	sccm
B	Helium flow	55	45	sccm
C	Power	220	180	watts
D	Pressure	220	180	mTorr

The unit of measure for the response was Å Oxide/minute.

ENCODED PARAMETERS (A,B,C,D)	RESPONSE OF REPLICATE 1	2	CONTRASTS
(– 1,– 1,– 1,– 1)	79.13	111.53	4011.80
(+1, – 1,– 1,– 1)	100.87	100.73	– 88.46
(– 1,+1, – 1,– 1)	96.87	118.73	235.22
(+1,+1, – 1,– 1)	118.40	107.47	– 3.20
(– 1,– 1,+1, – 1)	166.53	165.20	1086.60
(+1, – 1,+1, – 1)	172.13	173.20	– 92.34
(– 1,+1,+1, – 1)	193.33	174.27	45.74
(+1,+1,+1, – 1)	173.87	180.60	15.60
(– 1,– 1,– 1,+1)	67.27	81.20	– 453.92
(+1, – 1,– 1,+1)	75.80	67.40	– 131.82
(1,+1, – 1,+1)	94.00	81.60	46.78
(+1,+1, – 1,+1)	78.07	83.53	51.60
(– 1,– 1,+1,+1)	150.20	144.40	– 44.20
(+1, – 1,+1,+1)	112.30	120.40	– 50.86
(– 1,+1,+1,+1)	169.20	156.67	54.14
(+1,+1,+1,+1)	152.57	144.33	67.72

EFFECT	SUM OF SQUARES	DOF	MEAN SQUARES	FSTAT	SIGNIFICANCE	ESTIMATE OF EFFECT
1	1.61e7	1	1.61e7	1.82e5	>99.9	250.74
A	244.54	1	244.54	2.7642	>75.0	- 5.53
B	1729.01	1	1729.01	19.5443	>99.9	14.70
AB	0.32	1	0.32	0.0036	<75.0	- 0.20
C	36896.86	1	36896.86	417.0728	>99.9	67.91
AC	266.46	1	266.46	3.0120	>75.0	- 5.77
BC	65.38	1	65.38	0.7390	<75.0	2.86
ABC	7.60	1	7.60	0.0860	<75.0	0.97
D	6438.86	1	6438.86	72.7832	>99.0	- 28.37
AD	543.02	1	543.02	6.1381	>97.5	- 8.24
BD	68.39	1	68.39	0.7730	<75.0	2.92
ABD	83.20	1	83.20	0.9405	<75.0	3.22
CD	61.05	1	61.05	0.6901	<75.0	- 2.76
ACD	80.84	1	80.84	0.9137	<75.0	- 3.18
BCD	91.60	1	91.60	1.0354	<75.0	3.38
ABCD	143.31	1	143.31	1.6200	>75.0	4.23
ERROR	1415.46	16	88.47			
TOTAL	48135.89	31				

APPENDIX 6.1c. The Parameter Significance to the Photoresist Etch Rate

The parameters were evaluated in a full factorial experimental design.

PARAMETER	DESCRIPTION	HIGH VALUE	LOW VALUE	UNIT OF MEASURE
A	Chlorine flow	55	45	sccm
B	Helium flow	55	45	sccm
C	Power	220	180	watts
D	Pressure	220	180	mTorr

The unit of measure of the response was Å photoresist/minute.

ENCODED PARAMETERS (A,B,C,D)	RESPONSE OF REPLICATE 1	2	CONTRASTS
(– 1,– 1,– 1,– 1)	2454.18	2502.30	88428.11
(+1, – 1,– 1,– 1)	2437.30	2511.43	– 2770.05
(– 1,+1, – 1,– 1)	2537.00	2735.00	1981.29
(+1,+1, – 1,– 1)	2209.80	2416.50	429.05
(– 1,– 1,+1, – 1)	3329.10	3424.70	15865.09
(+1, – 1,+1, – 1)	3123.00	3346.90	1533.65
(– 1,+1,+1, – 1)	3308.00	3299.60	289.11
(+1,+1,+1, – 1)	3380.50	3176.20	1172.55
(– 1,– 1,– 1,+1)	2075.70	2271.70	– 3954.91
(+1, – 1,– 1,+1)	1631.80	1833.30	– 793.55
(– 1,+1, – 1,+1)	2247.90	2392.90	2113.91
(+1,+1, – 1,+1)	2072.70	1952.00	1238.95
(– 1,– 1,+1,+1)	3053.80	3300.00	2696.11
(+1, – 1,+1,+1)	2844.00	3084.20	896.35
(– 1,+1,+1,+1)	3385.90	3281.30	394.09
(+1,+1,+1,+1)	3607.80	3201.60	– 569.35

EFFECT	SUM OF SQUARES	DOF	MEAN SQUARES	FSTAT	SIGNIFICANCE	ESTIMATE OF EFFECT
1	7.82e9	1	7.82e9	2.61e+4	>99.9	5526.76
A	239786.78	1	239786.78	12.7954	<99.0	- 173.13
B	122672.19	1	122672.19	6.5460	>97.5	123.83
AB	5752.62	1	5752.62	0.3070	<75.0	26.82
C	7865658.77	1	7865658.77	419.7249	>99.9	991.57
AC	73502.57	1	73502.57	3.9222	>90.0	95.85
BC	2612.02	1	2612.02	0.1394	<75.0	18.07
ABC	42964.80	1	42964.80	2.2927	>75.0	73.28
D	488791.03	1	488791.03	26.0827	>99.9	- 247.18
AD	19678.80	1	19678.80	1.0501	<75.0	- 49.60
BD	139644.23	1	139644.23	7.4517	>97.5	132.12
ABD	47968.66	1	47968.66	2.5597	>75.0	77.43
CD	227156.54	1	227156.54	12.1215	>99.0	168.51
ACD	25107.60	1	25107.60	1.3398	<75.0	56.02
BCD	4853.34	1	4853.34	0.2590	<75.0	24.63
ABCD	10129.98	1	10129.98	0.5406	<75.0	- 35.58
ERROR	299840.56	16	18740.04			
TOTAL	9616120.50	31				

APPENDIX 6.1d. The Parameter Significance to the Polysilicon/Oxide Selectivity.

The parameters were evaluated in a full factorial experimental design.

PARAMETER	DESCRIPTION	HIGH VALUE	LOW VALUE	UNIT OF MEASURE
A	Chlorine flow	55	45	sccm
B	Helium flow	55	45	sccm
C	Power	220	180	watts
D	Pressure	220	180	mTorr

The unit of measure for the response was Å Polysilicon/Å Oxide.

ENCODED PARAMETERS (A,B,C,D)	RESPONSE OF REPLICATE 1	2	3	4	CONTRASTS
(– 1,– 1,– 1,– 1)	28.16	19.98	32.86	23.31	1538.31
(+1, – 1,– 1,– 1)	25.63	25.66	25.49	25.53	21.62
(– 1,+1, – 1,– 1)	26.48	21.60	27.80	22.69	– 67.47
(+1,+1, – 1,– 1)	20.52	22.61	22.51	24.80	12.53
(– 1,– 1,+1, – 1)	19.63	19.79	19.94	20.10	– 224.17
(+1, – 1,+1, – 1)	18.44	18.33	20.01	19.89	28.49
(– 1,+1,+1, – 1)	17.02	18.88	14.68	16.28	– 6.94
(+1,+1,+1, – 1)	18.94	18.24	19.06	18.35	– 0.95
(– 1,– 1,– 1,+1)	33.60	27.83	35.38	29.31	151.91
(+1, – 1,– 1,+1)	28.47	32.02	29.43	33.10	32.02
(– 1,+1, – 1,+1)	26.32	30.32	26.35	30.36	– 2.90
(+1,+1, – 1,+1)	30.82	28.80	33.15	30.36	3.80
(– 1,– 1,+1,+1)	21.63	22.50	22.14	23.02	– 28.06
(+1, – 1,+1,+1)	25.40	23.69	27.23	25.40	– 1.60
(– 1,+1,+1,+1)	19.92	21.51	18.73	20.23	– 12.77
(+1,+1,+1,+1)	21.41	22.64	21.41	22.64	– 34.25

EFFECT	SUM OF SQUARES	DOF	MEAN SQUARES	F_{STAT}	SIGNIFICANCE	ESTIMATE OF EFFECT
1	2.37e6	1	2.37e6	4.87e5	>99.9	48.07
A	7.30	1	7.30	1.5029	>75.0	0.68
B	71.13	1	71.13	14.6402	>99.0	- 2.11
AB	2.45	1	2.45	0.5045	<75.0	0.39
C	785.17	1	785.17	161.6003	>99.9	- 7.01
AC	12.68	1	12.68	2.6106	>75.0	0.89
BC	0.75	1	0.75	0.1547	<75.0	- 0.22
ABC	0.01	1	0.01	0.0029	<75.0	- 0.03
D	360.56	1	360.56	74.2095	>99.9	4.75
AD	16.02	1	16.02	3.2962	>90.0	1.00
BD	0.13	1	0.13	0.0271	<75.0	- 0.09
ABD	0.23	1	0.23	0.0464	<75.0	0.12
CD	12.30	1	12.30	2.5322	>75.0	- 0.88
ACD	0.04	1	0.04	0.0083	<75.0	- 0.05
BCD	2.55	1	2.55	0.5243	<75.0	- 0.40
ABCD	18.32	1	18.32	3.7714	>90.0	- 1.07
ERROR	233.22	48	4.86			
TOTAL	1522.88	63				

APPENDIX 6.1e. The Parameter Significance to the Polysilicon/ Photoresist Selectivity.

The parameters were evaluated in a full factorial experimental design.

PARAMETER	DESCRIPTION	HIGH VALUE	LOW VALUE	UNIT OF MEASURE
A	Chlorine flow	55	45	sccm
B	Helium flow	55	45	sccm
C	Power	220	180	watts
D	Pressure	220	180	mTorr

The unit of measure of the response was Å Polysilicon/Å photoresist.

ENCODED PARAMETERS (A,B,C,D)	RESPONSE OF REPLICATE 1	2	3	4	CONTRASTS
(– 1,– 1,– 1,– 1)	0.91	0.96	1.04	1.04	66.76
(+1. – 1,– 1,– 1)	1.06	1.03	1.06	1.02	2.46
(– 1,+1, – 1,– 1)	1.01	0.94	1.06	0.95	– 0.47
(+1,+1, – 1,– 1)	1.10	1.01	1.21	1.10	0.31
(– 1,– 1,+1, – 1)	0.98	0.95	1.00	0.97	– 3.44
(+1, – 1,+1, – 1)	1.02	0.95	1.10	1.03	– 1.87
(– 1,+1,+1, – 1)	0.99	1.00	0.86	0.86	– 0.77
(+1,+1,+1, – 1)	0.97	1.04	0.98	1.04	0.10
(– 1,– 1,– 1,+1)	1.09	0.99	1.15	1.05	2.30
(+1, – 1,– 1,+1)	1.32	1.18	1.37	1.22	0.07
(– 1,+1, – 1,+1)	1.10	1.03	1.10	1.05	– 0.47
(+1,+1, – 1,+1)	1.16	1.23	1.25	1.33	– 0.41
(– 1,– 1,+1,+1)	1.06	0.98	1.09	1.01	– 1.96
(+1, – 1,+1,+1)	1.00	0.92	1.08	0.99	– 1.55
(– 1,+1,+1,+1)	1.00	1.03	0.94	0.97	0.26
(+1,+1,+1,+1)	0.91	1.02	0.91	1.02	0.30

EFFECT	SUM OF SQUARES	DOF	MEAN SQUARES	FSTAT	SIGNIFICANCE	ESTIMATE OF EFFECT
1	69.64	1	69.64	1.97e4	>99.9	2.09
A	9.46e-2	1	9.46e-2	26.8252	>99.0	0.08
B	3.45e-3	1	3.45e-3	0.9646	<75.0	-0.01
AB	1.50e-3	1	1.50e-3	0.4385	<75.0	0.01
C	1.85e-1	1	1.85e-1	52.5334	>99.9	-0.11
AC	5.46e-2	1	5.46e-2	15.4966	>99.0	-0.06
BC	9.26e-3	1	9.26e-3	2.6365	>75.0	-0.02
ABC	1.57e-4	1	1.57e-4	0.0466	<75.0	0.00
D	8.27e-2	1	8.27e-2	23.4541	>99.9	0.07
AD	7.66e-5	1	7.66e-5	0.0193	<75.0	0.00
BD	3.45e-3	1	3.45e-3	0.9871	<75.0	-0.01
ABD	2.63e-3	1	2.63e-3	0.7448	<75.0	-0.01
CD	6.00e-2	1	6.00e-2	17.0205	>99.9	-0.06
ACD	3.75e-2	1	3.75e-2	10.6115	>99.0	-0.05
BCD	2.11e-3	1	2.11e-3	0.2967	<75.0	0.01
ABCD	1.41e-3	1	1.41e-3	0.4073	<75.0	0.01
ERROR	0.17	48	3.56e-3			
TOTAL	0.71	63				

APPENDIX 6.1f. The Parameter Significance to the Anisotropy.

The parameters were evaluated in a full factorial experimental design.

PARAMETER	DESCRIPTION	HIGH VALUE	LOW VALUE	UNIT OF MEASURE
A	Chlorine flow	55	45	sccm
B	Helium flow	55	45	sccm
C	Power	220	180	watts
D	Pressure	220	180	mTorr

The unit of measure for the response was a ratio and was, therefore, dimensionless.

ENCODED PARAMETERS (A,B,C,D)	RESPONSE OF REPLICATE 1	2	CONTRASTS
(– 1,– 1,– 1,– 1)	0.33	0.57	17.35
(+1, – 1,– 1,– 1)	0.55	0.60	1.04
(– 1,+1, – 1,– 1)	0.70	0.68	– 0.05
(+1,+1, – 1,– 1)	0.57	0.68	– 0.35
(– 1,– 1,+1, – 1)	0.42	0.35	– 2.56
(+1, – 1,+1, – 1)	0.51	0.51	– 0.26
(– 1,+1,+1, – 1)	0.35	0.36	– 1.55
(+1,+1,+1, – 1)	0.52	0.49	0.55
(– 1,– 1,– 1,+1)	0.62	0.51	0.95
(+1, – 1,– 1,+1)	0.72	0.71	– 0.31
(– 1,+1, – 1,+1)	0.60	0.65	– 1.12
(+1,+1, – 1,+1)	0.72	0.75	0.29
(– 1,– 1,+1,+1)	0.53	0.68	– 0.23
(+1, – 1,+1,+1)	0.60	0.51	– 1.13
(– 1,+1,+1,+1)	0.47	0.34	– 0.26
(+1,+1,+1,+1)	0.31	0.44	– 0.29

EFFECT	SUM OF SQUARES	DOF	MEAN SQUARES	F_{STAT}	SIGNIFICANCE	ESTIMATE OF EFFECT
1	9.41	1	9.41	117.59	>99.9	1.08
A	0.03	1	0.03	6.7725	>97.5	0.07
B	0.00	1	0.00	0.0145	<75.0	- 0.00
AB	0.00	1	0.00	0.7705	<75.0	- 0.02
C	0.20	1	0.20	40.7583	>99.0	- 0.16
AC	0.00	1	0.00	0.4355	<75.0	- 0.02
BC	0.08	1	0.08	14.9408	>99.0	- 0.10
ABC	0.01	1	0.01	1.8698	>75.0	0.03
D	0.03	1	0.03	5.6314	>95.0	0.06
AD	0.00	1	0.00	0.5969	<75.0	- 0.02
BD	0.04	1	0.04	7.7584	>97.5	- 0.07
ABD	0.00	1	0.00	0.5259	<75.0	0.02
CD	0.00	1	0.00	0.3400	<75.0	- 0.01
ACD	0.04	1	0.04	7.9653	>97.5	- 0.07
BCD	0.00	1	0.00	0.4185	<75.0	- 0.02
ABCD	0.00	1	0.00	0.5346	<75.0	- 0.02
ERROR	0.08	16	0.01			
TOTAL	0.53	31				

APPENDIX 6.2a. The First-Order Model of the Polysilicon Etch Rate.

The first-order model was derived from a 2^k factorial experiment.

PARAMETER	DESCRIPTION	HIGH VALUE	LOW VALUE	UNIT OF MEASURE
A	Chlorine flow	55	45	sccm
B	Helium flow	55	45	sccm
C	Power	220	180	watts
D	Pressure	220	180	mTorr

The response unit of measure was Å polysilicon/minute.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2
- 1.000	- 1.000	- 1.000	- 1.000	2228.00	2600.00
+1.000	- 1.000	- 1.000	- 1.000	2585.00	2571.50
- 1.000	+1.000	- 1.000	- 1.000	2565.00	2693.45
+1.000	+1.000	- 1.000	- 1.000	2430.00	2665.00
- 1.000	- 1.000	+1.000	- 1.000	3269.00	3320.00
+1.000	- 1.000	+1.000	- 1.000	3174.00	3445.00
- 1.000	+1.000	+1.000	- 1.000	3291.00	2837.40
+1.000	+1.000	+1.000	- 1.000	3293.40	3313.27
- 1.000	- 1.000	- 1.000	+1.000	2260.00	2380.00
+1.000	- 1.000	- 1.000	+1.000	2158.00	2231.00
- 1.000	+1.000	- 1.000	+1.000	2474.00	2477.00
+1.000	+1.000	- 1.000	+1.000	2406.00	2588.00
- 1.000	- 1.000	+1.000	+1.000	3249.00	3324.80
+1.000	- 1 000	+1.000	+1.000	2852.00	3058.20
- 1.000	+1.000	+1.000	+1.000	3370.00	3169.60
+1.000	+1.000	+1.000	+1.000	3267.00	3267.00

The first order model was written as

$$y = b_0 + \mathbf{x'b}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 2838$$

$$\mathbf{b'} = (-6.371 \quad 43.80 \quad 380.9 \quad -54.67).$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	2.577e+08	1	2.577e+08	1.310e+04
R| β_0	4.801e+ 06	4	1.200e+ 06	6.060e+ 01
LACK OF FIT	3.779e+05	11	3,436e+04	1.735e+00
PURE ERROR	3.169e+05	16	1.981e+04	
TOTAL	2.632e+08	32		

APPENDIX 6.2b. The First-Order Model of the Oxide Etch Rate.

The first-order model was derived from a 2^k factorial experiment.

PARAMETER	DESCRIPTION	HIGH VALUE	LOW VALUE	UNIT OF MEASURE
A	Chlorine flow	55	45	sccm
B	Helium flow	55	45	sccm
C	Power	220	180	watts
D	Pressure	220	180	mTorr

The unit of measure of the response was Å oxide/minute.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2
– 1.000	– 1.000	– 1.000	– 1.000	79.13	111.53
+1.000	– 1.000	– 1.000	– 1.000	100.87	100.73
– 1.000	+1.000	– 1.000	– 1.000	96.87	118.73
+1.000	+1.000	– 1.000	– 1.000	118.40	107.47
– 1.000	– 1.000	+1.000	– 1.000	166.53	165.20
+1.000	– 1.000	+1.000	– 1.000	172.13	173.20
– 1.000	+1.000	+1.000	– 1.000	193.33	174.27
+1.000	+1.000	+1.000	– 1.000	173.87	180.60
– 1.000	– 1.000	– 1.000	+1.000	67.27	81.20
+1.000	– 1.000	– 1.000	+1.000	75.80	67.40
– 1.000	+1.000	– 1.000	+1.000	94.00	81.60
+1.000	+1.000	– 1.000	+1.000	78.07	83.53
– 1.000	– 1.000	+1.000	+1.000	150.20	144.40
+1.000	– 1.000	+1.000	+1.000	112.30	120.40
– 1.000	+1.000	+1.000	+1.000	169.20	156.67
+1.000	+1.000	+1.000	+1.000	152.57	144.33

APPENDIX 6.2d. The First-Order Model of the Polysilicon/ Oxide Selectivity.

The first-order model was derived from a 2^k factorial experiment.

PARAMETER	DESCRIPTION	HIGH VALUE	LOW VALUE	UNIT OF MEASURE
A	Chlorine flow	55	45	sccm
B	Helium flow	55	45	sccm
C	Power	220	180	watts
D	Pressure	220	180	mTorr

The response unit of measure was Å polysilicon/Å Oxide.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2	3	4
− 1.000	− 1.000	− 1.000	− 1.000	28.16	19.98	32.86	23.31
+1.000	− 1.000	− 1.000	− 1.000	25.63	25.66	25.49	25.53
− 1.000	+1.000	− 1.000	− 1.000	26.48	21.60	27.80	22.69
+1.000	+1.000	− 1.000	− 1.000	20.52	22.61	22.51	24.80
− 1.000	− 1.000	+1.000	− 1.000	19.63	19.79	19.94	20.10
+1.000	− 1.000	+1.000	− 1.000	18.44	18.33	20.01	19.89
− 1.000	+1.000	+1.000	− 1.000	17.02	18.88	14.68	16.28
+1.000	+1.000	+1.000	− 1.000	18.94	18.24	19.06	18.35
− 1.000	− 1.000	− 1.000	+1.000	33.60	27.83	35.38	29.31
+1.000	− 1.000	− 1.000	+1.000	28.47	32.02	29.43	33.10
− 1.000	+1.000	− 1.000	+1.000	26.32	30.32	26.35	30.36
+1.000	+1.000	− 1.000	+1.000	30.82	28.80	33.15	30.36
− 1.000	− 1.000	+1.000	+1.000	21.63	22.50	22.14	23.02
+1.000	− 1.000	+1.000	+1.000	25.40	23.69	27.23	25.40
− 1.000	+1.000	+1.000	+1.000	19.92	21.51	18.73	20.23
+1.000	+1.000	+1.000	+1.000	21.41	22.64	21.41	22.64

The first order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 2838,$$

$$\mathbf{b}' = (3.378e-01 \quad -1.540e+00 \quad -3.503e+00 \quad 2.374e+00).$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	3.697e+04	1	3.697e+04	7.610e+03
R | β_0	1.224e+03	4	3.060e+02	6.299e+01
LACK OF FIT	6.549e+01	11	5.953e+00	1.225e+00
PURE ERROR	2.332e+02	48	4.859e+00	
TOTAL	3.850e+04	64		

APPENDIX 6.2e. The First-Order Model of the Polysilicon/Photoresist Selectivity.

The first-order model was derived from a 2^k factorial experiment.

PARAMETER	DESCRIPTION	HIGH VALUE	LOW VALUE	UNIT OF MEASURE
A	Chlorine flow	55	45	sccm
B	Helium flow	55	45	sccm
C	Power	220	180	watts
D	Pressure	220	180	mTorr

The unit of measure of the response was Å polysilicon/Å photoresist.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2	3	4
− 1.000	− 1.000	− 1.000	− 1.000	0.91	0.96	1.04	1.04
+1.000	− 1.000	− 1.000	− 1.000	1.06	1.03	1.06	1.02
− 1.000	+1.000	− 1.000	− 1.000	1.01	0.94	1.06	0.95
+1.000	+1.000	− 1.000	− 1.000	1.10	1.01	1.21	1.10
− 1.000	− 1.000	+1.000	− 1.000	0.98	0.95	1.00	0.97
+1.000	− 1.000	+1.000	− 1.000	1.02	0.95	1.10	1.03
− 1.000	+1.000	+1.000	− 1.000	0.99	1.00	0.86	0.86
+1.000	+1.000	+1.000	− 1.000	0.97	1.04	0.98	1.04
− 1.000	− 1.000	− 1.000	+1.000	1.09	0.99	1.15	1.05
+1.000	− 1.000	− 1.000	+1.000	1.32	1.18	1.37	1.22
− 1.000	+1.000	− 1.000	+1.000	1.10	1.03	1.10	1.05
+1.000	+1.000	− 1.000	+1.000	1.16	1.23	1.25	1.33
− 1.000	− 1.000	+1.000	+1.000	1.06	0.98	1.09	1.01
+1.000	− 1.000	+1.000	+1.000	1.00	0.92	1.08	0.99
− 1.000	+1.000	+1.000	+1.000	1.00	1.03	0.94	0.97
+1.000	+1.000	+1.000	+1.000	0.91	1.02	0.91	1.02

The first order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 1.043,$$

$$\mathbf{b}' = (3.845e-02 \quad -7.291e-03 \quad -5.380e-02 \quad 3.595e-02).$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	6.95e+01	1	6.95e+01	1.975e+04
R I β_0	3.660e-01	4	9.150e-02	2.594e+01
LACK OF FIT	1.718e-01	11	1.562e-02	4.428e+00
PURE ERROR	1.693e-01	48	3.527e-03	
TOTAL	7.035e+01	64		

APPENDIX 6.2f. The First-Order Model of the Anisotropy.

The first-order model was derived from a 2^k factorial experiment.

PARAMETER	DESCRIPTION	HIGH VALUE	LOW VALUE	UNIT OF MEASURE
A	Chlorine flow	55	45	sccm
B	Helium flow	55	45	sccm
C	Power	220	180	watts
D	Pressure	220	180	mTorr

The unit of measure of the response was 1 minus the ratio of the horizontal etch to the vertical etch.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2
− 1.000	− 1.000	− 1.000	− 1.000	0.33	0.57
+1.000	− 1.000	− 1.000	− 1.000	0.55	0.60
− 1.000	+1.000	− 1.000	− 1.000	0.70	0.68
+1.000	+1.000	− 1.000	− 1.000	0.57	0.68
− 1.000	− 1.000	+1.000	− 1.000	0.42	0.35
+1.000	− 1.000	+1.000	− 1.000	0.51	0.51
− 1.000	+1.000	+1.000	− 1.000	0.35	0.36
+1.000	+1.000	+1.000	− 1.000	0.52	0.49
− 1.000	− 1.000	− 1.000	+1.000	0.62	0.51
+1.000	− 1.000	− 1.000	+1.000	0.72	0.71
− 1.000	+1.000	− 1.000	+1.000	0.60	0.65
+1.000	+1.000	− 1.000	+1.000	0.72	0.75
− 1.000	− 1.000	+1.000	+1.000	0.53	0.68
+1.000	− 1.000	+1.000	+1.000	0.60	0.51
− 1.000	+1.000	+1.000	+1.000	0.47	0.34
+1.000	+1.000	+1.000	+1.000	0.31	0.44

The first order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 5.442e-01,$$

$$\mathbf{b}' = (3.262e-02 \quad -1.509e-03 \quad -8.003e-02 \quad 2.975e-02).$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	9.408e+00	1	9.408e+00	1.871e+03
R | β_0	2.674e-01	4	6.685e-02	1.329e+01
LACK OF FIT	1.818e-01	11	1.653e-02	3.287e+00
PURE ERROR	8.045e-02	16	5.028e-03	
TOTAL	9.938e+00	32		

APPENDIX 6.3a. The Second-Order Model of the Polysilicon Etch Rate

The second-order model was derived from a central composite experiment.

PARAMETER	DESCRIPTION	HIGH AXIAL VALUE	HIGH VALUE	CENTER VALUE	LOW VALUE	LOW AXIAL VALUE	UNIT OF MEASURE
A	Chlorine flow	57.1	55	50	45	42.9	sccm
B	Helium flow	57.1	55	50	45	42.9	sccm
C	Power	228.3	220	200	180	171.7	watts
D	Pressure	228.3	220	200	180	171.7	mTorr

The unit of measure of the response was Å polysilicon/minute.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2
− 1.000	− 1.000	− 1.000	− 1.000	2228.00	2600.00
+1.000	− 1.000	− 1.000	− 1.000	2585.00	2571.50
− 1.000	+1.000	− 1.000	− 1.000	2565.00	2693.45
+1.000	+1.000	− 1.000	− 1.000	2430.00	2665.00
− 1.000	− 1.000	+1.000	− 1.000	3269.00	3320.00
+1.000	− 1.000	+1.000	− 1.000	3174.00	3445.00
− 1.000	+1.000	+1.000	− 1.000	3291.00	2837.40
+1.000	+1.000	+1.000	− 1.000	3293.40	3313.27
− 1.000	− 1.000	− 1.000	+1.000	2260.00	2380.00
+1.000	− 1.000	− 1.000	+1.000	2158.00	2231.00
− 1.000	+1.000	− 1.000	+1.000	2474.00	2477.00
+1.000	+1.000	− 1.000	+1.000	2406.00	2588.00
− 1.000	− 1.000	+1.000	+1.000	3249.00	3324.80
+1.000	− 1.000	+1.000	+1.000	2852.00	3058.20
− 1.000	+1.000	+1.000	+1.000	3370.00	3169.60
+1.000	+1.000	+1.000	+1.000	3267.00	3267.00
+1.414	0.000	0.000	0.000	2679.00	
− 1.414	0.000	0.000	0.000	2952.00	
0.000	+1.414	0.000	0.000	2870.00	
0.000	− 1.414	0.000	0.000	2732.00	
0.000	0.000	+1.414	0.000	3194.00	
0.000	0.000	− 1.414	0.000	2133.00	
0.000	0.000	0.000	+1.414	2403.00	
0.000	0.000	0.000	− 1.414	2802.00	
0.000	0.000	0.000	0.000	2877.00	

The second-order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 2698$$

$$\mathbf{b}' = \begin{pmatrix} -1.639e+01 & 4.436e+01 & 3.803e+02 & -6.427e+01 \end{pmatrix},$$

$$\mathbf{B} = \begin{bmatrix} 8.096e+01 & 1.419e+01 & -1.843e+00 & -2.423e+01 \\ 1.419e+01 & 7.370e+01 & -1.825e+01 & 2.515e+01 \\ -1.843e+00 & -1.825e+01 & 4.932e+00 & 1.529e+01 \\ -2.423e+01 & 2.515e+01 & 1.529e+01 & -2.558e+01 \end{bmatrix}.$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	3.251e+08	1	3.251e+08	1.642e+04
R| β_0	5.798e+06	14	4.141e+05	2.091e+01
LACK OF FIT	2.276e+05	10	2.276e+04	1.149e+00
PURE ERROR	3.169e+05	16	1.981e+04	
TOTAL	3.315e+08	41		

STATIONARY POINT

$x_0 =$ (-2.702e+00 -6.737e-01 -1.544e+01 -8.587e+00)

CHARACTERISITIC ROOTS

$\lambda =$ (9.407e+01 7.555e+01 9.183e+00 -4.479e+01)

M MATRIX

$$M = \begin{bmatrix} 7.723e-01 & -5.906e-01 & 1.322e-01 & 1.930e-01 \\ 6.155e-01 & 7.377e-01 & 8.489e-02 & -2.639e-01 \\ -1.499e-01 & -1.085e-01 & 9.150e-01 & -3.586e-01 \\ -4.617e-02 & 3.086e-01 & 3.717e-01 & 8.743e-01 \end{bmatrix}.$$

CANONICAL ANALYSIS TABLE

	1	2	3	4
SLOPES	-39.404	-18.685	325.633	-207.430
CURVATURES	94.072	75.545	6.183	-44.787
AXIS DISTANCE	0.209	0.124	-17.729	-2.316
CHANGES IN y	116.222	82.186	564.087	362.060
FEATURE DISTANCE	17.880	17.881	2.328	17.731

APPENDIX 6.3b. The Second-Order Model of the Oxide Etch Rate

The second-order model was derived from a central composite experiment.

PARAMETER	DESCRIPTION	HIGH AXIAL VALUE	HIGH VALUE	CENTER VALUE	LOW VALUE	LOW AXIAL VALUE	UNIT OF MEASURE
A	Chlorine flow	57.1	55	50	45	42.9	sccm
B	Helium flow	57.1	55	50	45	42.9	sccm
C	Power	228.3	220	200	180	171.7	watts
D	Pressure	228.3	220	200	180	171.7	mTorr

The unit of measure of the response was Å oxide/minute.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2
− 1.000	− 1.000	− 1.000	− 1.000	79.13	111.53
+1.000	− 1.000	− 1.000	− 1.000	100.87	100.73
− 1.000	+1.000	− 1.000	− 1.000	96.87	118.7
+1.000	+1.000	− 1.000	− 1.000	118.40	107.47
− 1.000	− 1.000	+1.000	− 1.000	166.53	165.20
+1.000	− 1.000	+1.000	− 1.000	172.13	173.20
− 1.000	+1.000	+1.000	− 1.000	193.33	174.27
+1.000	+1.000	+1.000	− 1.000	173.87	180.60
− 1.000	− 1.000	− 1.000	+1.000	67.27	81.20
+1.000	− 1.000	− 1.000	+1.000	75.80	67.40
− 1.000	+1.000	− 1.000	+1.000	94.00	81.60
+1.000	+1.000	− 1.000	+1.000	78.07	83.53
− 1.000	− 1.000	+1.000	+1.000	150.20	144.40
+1.000	− 1.000	+1.000	+1.000	112.30	120.40
− 1.000	+1.000	+1.000	+1.000	169.20	156.67
+1.000	+1.000	+1.000	+1.000	152.57	144.33
+1.414	0.000	0.000	0.000	111.30	
− 1.414	0.000	0.000	0.000	126.42	
0.000	+1.414	0.000	0.000	126.25	
0.000	− 1.414	0.000	0.000	97.08	
0.000	0.000	+1.414	0.000	167.21	
0.000	0.000	− 1.414	0.000	53.25	
0.000	0.000	0.000	+1.414	79.80	
0.000	0.000	0.000	− 1.414	149.30	
0.000	0.000	0.000	0.000	83.70	

The second-order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 95.84$$

$$\mathbf{b}' = \begin{pmatrix} -3.051e+01 & 7.680e+00 & 3.466e+01 & -2.060e+01 \end{pmatrix},$$

$$\mathbf{B} = \begin{bmatrix} 9.994e+00 & -5.000e-02 & -1.443e+00 & -2.060e+00 \\ -5.000e-02 & 6.395e00 & 7.147e-01 & 7.309e-01 \\ -1.443e+00 & 7.147e-01 & 5.677e+00 & -6.906e-01 \\ -2.060e+00 & 7.309e-01 & -6.906e-01 & 7.838e+00 \end{bmatrix}.$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	6.112e+05	1	6.112e+05	6.909e+03
R I β_0	5.741e+ 04	14	4.100e+ 03	4.635e+ 01
LACK OF FIT	1.214e+03	10	1.214e+02	1.372e+00
PURE ERROR	1.415e+03	16	8.847e+01	
TOTAL	6.713e+05	41		

STATIONARY POINT

$x_0 =$ (-1.275e-01 -3.535e-01 -2.953e+00 7.178e-01)

CHARACTERISITIC ROOTS

$\lambda =$ (1.144e=01 7.509e+00 6.687e+00 4.272e+00)

M MATRIX

$$M = \begin{bmatrix} 8.558e\text{-}01 & 3.516e\text{-}01 & 1.761e\text{-}01 & 3.362e\text{-}01 \\ -1.019e\text{-}01 & 1.685e\text{-}01 & 9.002e\text{-}01 & -3.884e\text{-}01 \\ -1.697e\text{-}01 & -5.023e\text{-}01 & 3.979e\text{-}01 & 7.488e\text{-}01 \\ -4.780e\text{-}01 & 7.719e\text{-}01 & -1.781e\text{-}02 & 4.189e\text{-}01 \end{bmatrix}.$$

CANONICAL ANALYSIS TABLE

	1	2	3	4
SLOPES	-1.944	-29.027	20.439	15.518
CURVATURES	11.436	7.509	6.687	4.272
AXIS DISTANCE	0.085	1.1933	-1.528	-1.816
CHANGES IN y	11.922	50.834	36.028	27.216
FEATURE DISTANCE	3.061	2.375	2.654	2.466

APPENDIX 6.3c. The Second-Order Model of the Photoresist Etch Rate

The second-order model was derived from a central composite experiment.

PARAMETER	DESCRIPTION	HIGH AXIAL VALUE	HIGH VALUE	CENTER VALUE	LOW VALUE	LOW AXIAL VALUE	UNIT OF MEASURE
A	Chlorine flow	57.1	55	50	45	42.9	sccm
B	Helium flow	57.1	55	50	45	42.9	sccm
C	Power	228.3	220	200	180	171.7	watts
D	Pressure	228.3	220	200	180	171.7	mTorr

The unit of measure of the response was Å photoresist/minute.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2
−1.000	−1.000	−1.000	−1.000	2454.18	2502.30
+1.000	−1.000	−1.000	−1.000	2437.30	2511.43
−1.000	+1.000	−1.000	−1.000	2537.00	2735.00
+1.000	+1.000	−1.000	−1.000	2209.80	2416.50
−1.000	−1.000	+1.000	−1.000	3329.10	3424.70
+1.000	−1.000	+1.000	−1.000	3123.00	3346.90
−1.000	+1.000	+1.000	−1.000	3308.00	3299.60
+1.000	+1.000	+1.000	−1.000	3380.50	3176.20
−1.000	−1.000	−1.000	+1.000	2075.70	2271.70
+1.000	−1.000	−1.000	+1.000	1631.80	1833.30
−1.000	+1.000	−1.000	+1.000	2247.90	2392.90
+1.000	+1.000	−1.000	+1.000	2072.70	1952.00
−1.000	−1.000	+1.000	+1.000	3053.80	3300.00
+1.000	−1.000	+1.000	+1.000	2844.00	3084.20
−1.000	+1.000	+1.000	+1.000	3385.90	3281.30
+1.000	+1.000	+1.000	+1.000	3607.80	3201.60
+1.414	0.000	0.000	0.000	2579.00	
−1.414	0.000	0.000	0.000	2916.00	
0.000	+1.414	0.000	0.000	2579.00	
0.000	−1.414	0.000	0.000	2814.00	
0.000	0.000	+1.414	0.000	3100.00	
0.000	0.000	−1.414	0.000	1808.00	
0.000	0.000	0.000	+1.414	2156.00	
0.000	0.000	0.000	−1.414	2415.00	
0.000	0.000	0.000	0.000	2655.00	

The second-order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 2442,$$

$$\mathbf{b}' = \begin{pmatrix} -9.019e01 & 4.581e+01 & 4.925e+02 & -1.200e+02 \end{pmatrix},$$

$$\mathbf{B} = \begin{bmatrix} 1.796e+02 & 6.704e+00 & 2.396e+01 & -1.240e+01 \\ 6.704e+00 & 1.541e+02 & 4.517e+00 & 3.303e+01 \\ 2.396e+01 & 4.517e+00 & 3.280e+01 & 4.213e+01 \\ -1.240e+01 & 3.303e+01 & 4.213e+01 & -5.148e+01 \end{bmatrix}.$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	3.030e+08	1	3.030e+08	1.617e+04
R | β_0	1.057e+07	14	7.548e+05	4.028e+01
LACK OF FIT	2.882e+05	10	2.882e+04	1.538e+00
PURE ERROR	2.998e+05	16	1.874e+04	
TOTAL	3.141e+08	41		

STATIONARY POINT

$x_0 =$ (4.249e-01 6.923e-01 -3.335e+00 -3.553e+00).

CHARACTERISITIC ROOTS

$\lambda =$ (1.852e+02 1.585e+02 4.565e+01 -7.437e+01).

M MATRIX

$$M = \begin{bmatrix} 9.565e-01 & -2.513e-01 & -1.266e-01 & 8.325e-02 \\ 2.429e-01 & 9.512e-01 & -1.412e-01 & -1.274e-01 \\ 1.611e-01 & 4.435e-02 & 9.127e-01 & -3.730e-01 \\ 1.247e-02 & 1.733e-01 & 3.635e-01 & 9.163e-01 \end{bmatrix}.$$

CANONICAL ANALYSIS TABLE

	1	2	3	4
SLOPES	2.533	67.227	409.496	-306.530
CURVATURES	185.166	158.540	45.655	-74.374
AXIS DISTANCE	-0.007	-0.212	-4.485	-2.061
CHANGES IN y	185.218	196.706	710.735	536.110
FEATURE DISTANCE	4.940	4.935	2.072	4.490

APPENDIX 6.3d. The Second-Order Model of the Polysilicon/ Oxide Selectivity.

The second-order model was derived from a central composite experiment.

PARAMETER	DESCRIPTION	HIGH AXIAL VALUE	HIGH VALUE	CENTER VALUE	LOW VALUE	LOW AXIAL VALUE	UNIT OF MEASURE
A	Chlorine flow	57.1	55	50	45	42.9	sccm
B	Helium flow	57.1	55	50	45	42.9	sccm
C	Power	228.3	220	200	180	171.7	watts
D	Pressure	228.3	220	200	180	171.7	mTorr

The response unit of measure was Å polysilicon/Å oxide.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2	3	4
− 1.000	− 1.000	− 1.000	− 1.000	28.16	19.98	32.86	23.31
+1.000	− 1.000	− 1.000	− 1.000	25.63	25.66	25.49	25.53
− 1.000	+1.000	− 1.000	− 1.000	26.48	21.60	27.80	22.69
+1.000	+1.000	− 1.000	− 1.000	20.52	22.61	22.51	24.80
− 1.000	− 1.000	+1.000	− 1.000	19.63	19.79	19.94	20.10
+1.000	− 1.000	+1.000	− 1.000	18.44	18.33	20.01	19.89
− 1.000	+1.000	+1.000	− 1.000	17.02	18.88	14.68	16.28
+1.000	+1.000	+1.000	− 1.000	18.94	18.24	19.06	18.35
− 1.000	− 1.000	− 1.000	+1.000	33.60	27.83	35.38	29.31
+1.000	− 1.000	− 1.000	+1.000	28.47	32.02	29.43	33.10
− 1.000	+1.000	− 1.000	+1.000	26.32	30.32	26.35	30.36
+1.000	+1.000	− 1.000	+1.000	30.82	28.80	33.15	30.36
− 1.000	− 1.000	+1.000	+1.000	21.63	22.50	22.14	23.02
+1.000	− 1.000	+1.000	+1.000	25.40	23.69	27.23	25.40
− 1.000	+1.000	+1.000	+1.000	19.92	21.51	18.73	20.23
+1.000	+1.000	+1.000	+1.000	21.41	22.64	21.41	22.64
+1.414	0.000	0.000	0.000	34.37			
− 1.414	0.000	0.000	0.000	24.07			
0.000	+1.414	0.000	0.000	23.35			
0.000	− 1.414	0.000	0.000	22.73			
0.000	0.000	+1.414	0.000	28.14			
0.000	0.000	− 1.414	0.000	19.10			
0.000	0.000	0.000	+1.414	40.06			
0.000	0.000	0.000	− 1.414	30.11			
0.000	0.000	0.000	0.000	18.81			

The second-order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 29.87,$$

$$\mathbf{b}' = \begin{pmatrix} 3.329e-01 & -1.105e+00 & -3.732e+00 & 2.469e+00 \end{pmatrix},$$

$$\mathbf{B} = \begin{bmatrix} -2.520e+00 & 9.785e-02 & 2.226e-01 & 2.501e-01 \\ 9.785e-02 & -1.657e+00 & -5.418e-02 & -2.268e-02 \\ 2.226e-01 & -5.418e-02 & 4.157e-01 & -2.192e-01 \\ 2.501e-01 & -2.268e-02 & -2.193e-01 & -2.145e+00 \end{bmatrix}.$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	4.336e+04	1	4.336e+04	8.923e+03
R| β_0	1.613e+03	14	1.165e+02	2.397e+01
LACK OF FIT	1.220e+02	10	1.220e+01	2.511e+00
PURE ERROR	2.332e+02	48	4.859e+00	
TOTAL	4.534e+04	73		

STATIONARY POINT

$x_0 =$ (4.470e-01 -4.499e-01 4.293e+00 1.936e-01).

CHARACTERISITIC ROOTS

$\lambda =$ (4.490e-01 -1.647e+00 -2.025e+00 -2.684e+00).

M MATRIX

$$M = \begin{bmatrix} 6.732e-02 & 1.170e-01 & 4.535e-01 & 8.810e-01 \\ -2.162e-02 & 9.930e-01 & -6.044e-02 & -9.907e-02 \\ 9.945e-01 & 1.423e-02 & 3.710e-02 & -9.699e-02 \\ -7.736e-02 & 7.253e-03 & 8.884e-01 & -4.524e-01 \end{bmatrix}.$$

CANONICAL ANALYSIS TABLE

	1	2	3	4
SLOPES	-3.857	-1.093	2.273	-0.352
CURVATURES	0.449	-1.647	-2.025	-2.684
AXIS DISTANCE	4.295	-0.332	0.561	-0.066
CHANGES IN y	6.695	2.510	4.427	2.753
FEATURE DISTANCE	0.655	4.332	4.308	4.344

APPENDIX 6.3e. The Second-Order Model of the Polysilicon/ Photoresist Selectivity.

The second-order model was derived from a central composite experiment.

PARAMETER	DESCRIPTION	HIGH AXIAL VALUE	HIGH VALUE	CENTER VALUE	LOW VALUE	LOW AXIAL VALUE	UNIT OF MEASURE
A	Chlorine flow	57.1	55	50	45	42.9	sccm
B	Helium flow	57.1	55	50	45	42.9	sccm
C	Power	228.3	220	200	180	171.7	watts
D	Pressure	228.3	220	200	180	171.7	mTorr

The unit of measure of the response was Å polysilicon/Å photoresist.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2	3	4
− 1.000	− 1.000	− 1.000	− 1.000	0.91	0.96	1.04	1.04
+1.000	− 1.000	− 1.000	− 1.000	1.06	1.03	1.06	1.02
− 1.000	+1.000	− 1.000	− 1.000	1.01	0.94	1.06	0.95
+1.000	+1.000	− 1.000	− 1.000	1.10	1.01	1.21	1.10
− 1.000	− 1.000	+1.000	− 1.000	0.98	0.95	1.00	0.97
+1.000	− 1.000	+1.000	− 1.000	1.02	0.95	1.10	1.03
− 1.000	+1.000	+1.000	− 1.000	0.99	1.00	0.86	0.86
+1.000	+1.000	+1.000	− 1.000	0.97	1.04	0.98	1.04
− 1.000	− 1.000	− 1.000	+1.000	1.09	0.99	1.15	1.05
+1.000	− 1.000	− 1.000	+1.000	1.32	1.18	1.37	1.22
− 1.000	+1.000	− 1.000	+1.000	1.10	1.03	1.10	1.05
+1.000	+1.000	− 1.000	+1.000	1.16	1.23	1.25	1.33
− 1.000	− 1.000	+1.000	+1.000	1.06	0.98	1.09	1.01
+1.000	− 1.000	+1.000	+1.000	1.00	0.92	1.08	0.99
− 1.000	+1.000	+1.000	+1.000	1.00	1.03	0.94	0.97
+1.000	+1.000	+1.000	+1.000	0.91	1.02	0.91	1.02
+1.414	0.000	0.000	0.000	1.09			
− 1.414	0.000	0.000	0.000	1.01			
0.000	+1.414	0.000	0.000	1.11			
0.000	− 1.414	0.000	0.000	0.97			
0.000	0.000	+1.414	0.000	1.03			
0.000	0.000	− 1.414	0.000	1.18			
0.000	0.000	0.000	+1.414	1.11			
0.000	0.000	0.000	− 1.414	1.16			
0.000	0.000	0.000	0.000	1.08			

The second-order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 1.109,$$

$$\mathbf{b}' = \begin{pmatrix} 3.785e-01 & -3.913e-03 & -5.375e-02 & 3.289e-02 \end{pmatrix},$$

$$\mathbf{B} = \begin{bmatrix} -3.309e-02 & 2.458e-03 & -5.375e-02 & 5.156e-04 \\ 2.458e-03 & -3.715e-02 & -6.027e-03 & -3.687e-03 \\ -1.461e-02 & -6.027e-03 & -5.636e-03 & -1.531e-02 \\ 5.156e-04 & -3.687e-03 & -1.531e-02 & 1.057e-02 \end{bmatrix}.$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	8.019e+01	1	8.019e+01	2.274e+04
R| β_0	5.213e-01	14	3.724e-02	1.056e+01
LACK OF FIT	6.695e-02	10	6.695e-03	1.898e+00
PURE ERROR	1.693e-01	48	3.527e-03	
TOTAL	8.095e+01	73		

STATIONARY POINT

$x_0 =$ (8.638e-01 3.611e-01 -6.877e-01 -2.468e+00).

CHARACTERISITIC ROOTS

$\lambda =$ (2.107e-02 -7.291e-03 -3.854e-02 -4.055e-02).

M MATRIX

$$M = \begin{bmatrix} 1.586e-01 & -4.021e-01 & -5.131e-01 & 7.415e-01 \\ 1.271e-02 & -2.405e-01 & 8.553e-01 & 4.587e-01 \\ -5.569e-01 & 6.891e-01 & -4.539e-02 & 4.613e-01 \\ 8.152e-01 & 5.528e-01 & 5.546e-01 & 1.638e-01 \end{bmatrix}$$

CANONICAL ANALYSIS TABLE

	1	2	3	4
SLOPES	0.063	-0.033	-0.019	0.007
CURVATURES	0.021	-2.273	-0.240	0.085
AXIS DISTANCE	-1.488	-2.273	-0.240	0.085
CHANGES IN y	0.111	0.058	0.050	0.042
FEATURE DISTANCE	2.287	1.509	2.718	2.727

APPENDIX 6.3f. The Second-Order Model of the Anisotropy

The second-order model was derived from a central composite experiment.

PARAMETER	DESCRIPTION	HIGH AXIAL VALUE	HIGH VALUE	CENTER VALUE	LOW VALUE	LOW AXIAL VALUE	UNIT OF MEASURE
A	Chlorine flow	57.1	55	50	45	42.9	sccm
B	Helium flow	57.1	55	50	45	42.9	sccm
C	Power	228.3	220	200	180	171.7	watts
D	Pressure	228.3	220	200	180	171.7	mTorr

The unit of measure of anisotropy was dimensionless.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2
- 1.000	- 1.000	- 1.000	- 1.000	0.33	0.57
+1.000	- 1.000	- 1.000	- 1.000	0.55	0.60
- 1.000	+1.000	- 1.000	- 1.000	0.70	0.68
+1.000	+1.000	- 1.000	- 1.000	0.57	0.68
- 1.000	- 1.000	+1.000	- 1.000	0.42	0.35
+1.000	- 1.000	+1.000	- 1.000	0.51	0.51
- 1.000	+1.000	+1 000	- 1.000	0.35	0.36
+1.000	+1.000	+1.000	- 1.000	0.52	0.49
- 1.000	- 1.000	- 1.000	+1.000	0.62	0.51
+1.000	- 1.000	- 1.000	+1.000	0.72	0.71
- 1.000	+1.000	- 1.000	+1.000	0.60	0.65
+1.000	+1.000	- 1.000	+1.000	0.72	0.75
- 1.000	- 1.000	+1.000	+1.000	0.53	0.68
+1.000	- 1.000	+1.000	+1.000	0.60	0.51
- 1.000	+1.000	+1.000	+1.000	0.47	0.34
+1.000	+1.000	+1.000	+1.000	0.31	0.44
+1.414	0.000	0.000	0.000	0.13	
- 1.414	0.000	0.000	0.000	0.23	
0.000	+1.414	0.000	0.000	0.46	
0.000	- 1.414	0.000	0.000	0.51	
0.000	0.000	+1.414	0.000	- 0.11	
0.000	0.000	- 1.414	0.000	0.51	
0.000	0.000	0.000	+1.414	- 0.06	
0.000	0.000	0.000	- 1.414	0.40	
0.000	0.000	0.000	0.000	0.36	

The second-order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'B\mathbf{x}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 1.088e-01,$$

$$\mathbf{b}' = (\quad 2.507e-02 \quad -3.306e-03 \quad -9.549e-02 \quad 8.374e-03),$$

$$B = \begin{bmatrix} 6.701e-02 & -5.502e-03 & -4.136e-03 & -4.842e-03 \\ -5.502e-03 & 2.196e-01 & -2.423e-02 & -1.746e-02 \\ -4.136e-03 & -2.423e-02 & 7.701e-02 & -3.655e-03 \\ -4.842e-03 & -1.746e-02 & -3.655e-03 & 6201e-02 \end{bmatrix}$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	9.544e+00	1	9.544e+00	1.898e+03
R I β_0	1.049e+ 00	14	7.495e-02	1.490e+ 01
LACK OF FIT	3.713e-01	10	3.713e-02	7.385e+00
PURE ERROR	8.045e-02	16	5.028e-03	
TOTAL	1.104e+01	41		

STATIONARY POINT

$x_0 =$ (-1.435e-01 7.222e-02 6.340e-01 -2.103e-02).

CHARACTERISITIC ROOTS

$\lambda =$ (2.254e-01 7.657e-02 6.967e-02 5.397e-02).

M MATRIX

$$M = \begin{bmatrix} -2.694e-02 & -3.279e-01 & 8.172e-01 & 4.732e-01 \\ 9.821e-01 & 1.101e-01 & -1.180e-02 & 1.526e-01 \\ -1.571e-01 & 9.044e-01 & 1.436e-01 & 3.699e-01 \\ -1.006e-01 & -2.499e-01 & -5.581e-01 & 7.848e-01 \end{bmatrix}.$$

CANONICAL ANALYSIS TABLE

	1	2	3	4
SLOPES	0.010	-0.970	0.002	-0.017
CURVATURES	0.255	0.077	0.070	0.054
AXIS DISTANCE	-0.023	0.634	-0.015	0.161
CHANGES IN y	0.226	0.185	0.070	0.062
FEATURE DISTANCE	0.654	0.163	0.654	0.634

APPENDIX 6.3g. The Second-Order Model of the Polysilicon Etch Rate

The second-order model was derived from a central composite experiment.

PARAMETER	DESCRIPTION	HIGH AXIAL VALUE	HIGH VALUE	CENTER VALUE	LOW VALUE	LOW AXIAL VALUE	UNIT OF MEASURE
A	Chlorine flow	57.1	55	50	45	42.9	sccm
B	Helium flow	57.1	55	50	45	42.9	sccm
C	Power	228.3	220	200	180	171.7	watts
D	Pressure	228.3	220	200	180	171.7	mTorr

The unit of measure of the response was Å polysilicon/minute.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2
− 1.000	− 1.000	− 1.000	− 1.000	2228.00	2600.00
+1.000	− 1.000	− 1.000	− 1.000	2585.00	2571.50
− 1.000	+1.000	− 1.000	− 1.000	2565.00	2693.45
+1.000	+1.000	− 1.000	− 1.000	2430.00	2665.00
− 1.000	− 1.000	+1.000	− 1.000	3269.00	3320.00
+1.000	− 1.000	+1.000	− 1.000	3174.00	3445.00
− 1.000	+1.000	+1.000	− 1.000	3291.00	2837.40
+1.000	+1.000	+1.000	− 1.000	3293.40	3313.27
− 1.000	− 1.000	− 1.000	+1.000	2260.00	2380.00
+1.000	− 1.000	− 1.000	+1.000	2158.00	2231.00
− 1.000	+1.000	− 1.000	+1.000	2474.00	2477.00
+1.000	+1.000	− 1.000	+1.000	2406.00	2588.00
− 1.000	− 1.000	+1.000	+1.000	3249.00	3324.80
+1.000	− 1.000	+1.000	+1.000	2852.00	3058.20
− 1.000	+1.000	+1.000	+1.000	3370.00	3169.60
+1.000	+1.000	+1.000	+1.000	3267.00	3267.00
+1.414	0.000	0.025	0.050	2679.00	
− 1.414	0.000	0.025	0.050	2952.00	
0.000	+1.414	0.025	0.050	2870.00	
0.000	− 1.414	0.025	0.050	2732.00	
0.000	0.000	+1.391	0.050	3194.00	
0.000	0.000	− 1.441	0.050	2133.00	
0.000	0.000	0.025	+1.368	2403.00	
0.000	0.000	0.025	− 1.469	2802.00	
0.000	0.000	0.025	0.050	2877.00	

The second-order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 2689,$$

$$\mathbf{b}' = (-1.610e+01 \quad 4.418e+01 \quad 3.801e+02 \quad -6.460e+01),$$

$$\mathbf{B} = \begin{bmatrix} 8.069e+01 & 1.419e+01 & -1.965e+00 & -2.447e+01 \\ 1.419e+01 & 7.344e+01 & -1.825e+01 & 2.516e+01 \\ -1.965e+00 & -1.825e+01 & 1.431e+01 & 1.515e+01 \\ -2.447e+01 & 2.516e+01 & 1.515e+01 & -2.506e+01 \end{bmatrix}.$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	3.251e+08	1	3.251e+08	1.642e+04
R| β_0	5.797e+06	14	4.141e+05	2.091e+01
LACK OF FIT	2.279e+05	10	2.279e+04	1.151e+00
PURE ERROR	3.169e+05	16	1.981e+04	
TOTAL	3.315e+08	41		

STATIONARY POINT

$x_0 =$ (-1.613e+00 -3.258e-01 -8.463e+01 -5.156e+00).

CHARACTERISITIC ROOTS

$\lambda =$ (9.410e+01 7.557e+01 1.706e+01 -4.336e+01).

M MATRIX

$$M = \begin{bmatrix} 7.670e-01 & -5.956e-01 & 1.289e-01 & 2.010e-01 \\ 6.171e-01 & 7.312e-01 & 1.203e-02 & -2.648e-01 \\ -1.693e-01 & -1.222e-01 & 9.272e-01 & -3.109e-01 \\ -4.874e-02 & 3.092e-01 & 3.304e-01 & 8.904e-01 \end{bmatrix}.$$

CANONICAL ANALYSIS TABLE

	1	2	3	4
SLOPES	-46.284	-24.550	334.357	-190.617
CURVATURES	94.100	75.571	17.063	-43.359
AXIS DISTANCE	0.246	0.162	-9.797	-2.198
CHANGES IN y	123.618	86.713	579.375	332.994
FEATURE DISTANCE	10.042	10.044	2.218	9.802

APPENDIX 6.3h. The Second-Order Model of the Oxide Etch Rate

The second-order model was derived from a central composite experiment.

PARAMETER	DESCRIPTION	HIGH AXIAL VALUE	HIGH VALUE	CENTER VALUE	LOW VALUE	LOW AXIAL VALUE	UNIT OF MEASURE
A	Chlorine flow	57.1	55	50	45	42.9	sccm
B	Helium flow	57.1	55	50	45	42.9	sccm
C	Power	228.3	220	200	180	171.7	watts
D	Pressure	228.3	220	200	180	171.7	mTorr

The unit of measure of the response was Å oxide/minute.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2
− 1.000	− 1.000	− 1.000	− 1.000	79.13	111.53
+1.000	− 1.000	− 1.000	− 1.000	100.87	100.73
− 1.000	+1.000	− 1.000	− 1.000	96.87	118.7
+1.000	+1.000	− 1.000	− 1.000	118.40	107.47
− 1.000	− 1.000	+1.000	− 1.000	166.53	165.20
+1.000	− 1.000	+1.000	− 1.000	172.13	173.20
− 1.000	+1.000	+1.000	− 1.000	193.33	174.27
+1.000	+1.000	+1.000	− 1.000	173.87	180.60
− 1.000	− 1.000	− 1.000	+1.000	67.27	81.20
+1.000	− 1.000	− 1.000	+1.000	75.80	67.40
− 1.000	+1.000	− 1.000	+1.000	94.00	81.60
+1.000	+1.000	− 1.000	+1.000	78.07	83.53
− 1.000	− 1.000	+1.000	+1.000	150.20	144.40
+1.000	− 1.000	+1.000	+1.000	112.30	120.40
− 1.000	+1.000	+1.000	+1.000	169.20	156.67
+1.000	+1.000	+1.000	+1.000	152.57	144.33
+1.414	0.000	0.025	0.050	111.30	
− 1.414	0.000	0.025	0.050	126.42	
0.000	+1.414	0.025	0.050	126.25	
0.000	− 1.414	0.025	0.050	97.08	
0.000	0.000	+1.391	0.050	167.21	
0.000	0.000	− 1.441	0.050	53.25	
0.000	0.000	0.025	+1.368	79.80	
0.000	0.000	0.025	− 1.469	149.30	
0.000	0.000	0.025	0.050	83.70	

The second-order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 95.44,$$

$$\mathbf{b}' = \begin{pmatrix} -3.020e+01 & 7.668e+00 & 3.470e+01 & -1.524e+01 \end{pmatrix},$$

$$\mathbf{B} = \begin{bmatrix} 1.002e+01 & -5.000e-02 & -1.446e+00 & -2.066e+00 \\ -5.000e-02 & 6.420e00 & 7.187e-01 & 7.389e-01 \\ -1.446e+00 & 7.187e-01 & 6.411e+00 & -6.855e-01 \\ -2.066e+00 & 7.389e-01 & -6.855e-01 & 7.486e+00 \end{bmatrix}.$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	6.112e+05	1	6.112e+05	6.909e+03
R| β_0	5.746e+ 04	14	4.104e+ 03	4.639e+ 01
LACK OF FIT	1.163e+03	10	1.163e+02	1.314e+00
PURE ERROR	1.415e+03	16	8.847e+01	
TOTAL	6.713e+05	41		

STATIONARY POINT

$x_0 =$ (-5.902e-02 -4.004e-01 -2.589e+00 8.043e-01).

CHARACTERISITIC ROOTS

$\lambda =$ (1.141e+01 7.559e+00 6.809e+00 4.554e+00).

M MATRIX

$$M = \begin{bmatrix} 8.680e-01 & -2.080e-01 & 2.713e-01 & 3.602e-01 \\ -1.034e-01 & -3.685e-01 & 8.910e-01 & -4.409e-01 \\ -2.054e-01 & 6.667e-01 & 3.180e-01 & 6.410e-01 \\ -4.401e-01 & 7.141e-01 & -1.774e-02 & 5.147e-01 \end{bmatrix}.$$

CANONICAL ANALYSIS TABLE

	1	2	3	4
SLOPES	-3.832	34.429	14.343	9.928
CURVATURES	11.415	7.559	6.809	4.554
AXIS DISTANCE	0.168	-2.277	-1.053	-1.090
CHANGES IN y	13.204	60.110	25.759	17.788
FEATURE DISTANCE	2.736	1.525	2.530	2.515

APPENDIX 6.3i. The Second-Order Model of the Photoresist Etch Rate

The second-order model was derived from a central composite experiment.

PARAMETER	DESCRIPTION	HIGH AXIAL VALUE	HIGH VALUE	CENTER VALUE	LOW VALUE	LOW AXIAL VALUE	UNIT OF MEASURE
A	Chlorine flow	57.1	55	50	45	42.9	sccm
B	Helium flow	57.1	55	50	45	42.9	sccm
C	Power	228.3	220	200	180	171.7	watts
D	Pressure	228.3	220	200	180	171.7	mTorr

The unit of measure of the response was Å photoresist/minute.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2
− 1.000	− 1.000	− 1.000	− 1.000	2454.18	2502.30
+1.000	− 1.000	− 1.000	− 1.000	2437.30	2511.43
− 1.000	+1.000	− 1.000	− 1.000	2537.00	2735.00
+1.000	+1.000	− 1.000	− 1.000	2209.80	2416.50
− 1.000	− 1.000	+1.000	− 1.000	3329.10	3424.70
+1.000	− 1.000	+1.000	− 1.000	3123.00	3346.90
− 1.000	+1.000	+1.000	− 1.000	3308.00	3299.60
+1.000	+1.000	+1.000	− 1.000	3380.50	3176.20
− 1.000	− 1.000	− 1.000	+1.000	2075.70	2271.70
+1.000	− 1.000	− 1.000	+1.000	1631.80	1833.30
− 1.000	+1.000	− 1.000	+1.000	2247.90	2392.90
+1.000	+1.000	− 1.000	+1.000	2072.70	1952.00
− 1.000	− 1.000	+1.000	+1.000	3053.80	3300.00
+1.000	− 1.000	+1.000	+1.000	2844.00	3084.20
− 1.000	+1.000	+1.000	+1.000	3385.90	3281.30
+1.000	+1.000	+1.000	+1.000	3607.80	3201.60
+1.414	0.000	0.025	0.050	2579.00	
− 1.414	0.000	0.025	0.050	2916.00	
0.000	+1.414	0.025	0.050	2579.00	
0.000	− 1.414	0.025	0.050	2814.00	
0.000	0.000	+1.391	0.050	3100.00	
0.000	0.000	− 1.441	0.050	1808.00	
0.000	0.000	0.025	+1.368	2156.00	
0.000	0.000	0.025	− 1.469	2415.00	
0.000	0.000	0.025	0.050	2655.00	

The second-order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 2436,$$

$$\mathbf{b}' = \begin{pmatrix} -9.018e01 & 4.542e+01 & 4.912e+02 & -1.209e+02 \end{pmatrix},$$

$$\mathbf{B} = \begin{bmatrix} 1.794e+02 & 6.704e+00 & 2.392e+01 & -1.249e+01 \\ 6.704e+00 & 1.539e+02 & 4.311e+00 & 3.262e+01 \\ 2.392e+01 & 4.311e+00 & 4.577e+01 & 4.205e+01 \\ -1.249e+01 & 3.262e+01 & 4.205e+01 & -5.859e+01 \end{bmatrix}.$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	3.030e+08	1	3.030e+08	1.617e+04
R| β_0	1.056e+07	14	7.546e+05	4.027e+01
LACK OF FIT	2.908e+05	10	2.908e+04	1.552e+00
PURE ERROR	2.998e+05	16	1.874e+04	
TOTAL	3.141e+08	41		

STATIONARY POINT

$x_0 =$ (4.192e-01 5.381e-01 -2.941e+00 -2.933e+00).

CHARACTERISITIC ROOTS

$\lambda =$ (1.853e+02 1.580e+02 5.589e+01 -7.876e+01).

M MATRIX

$$M = \begin{bmatrix} 9.543e-01 & -2.520e-01 & -1.402e-01 & 7.866e-02 \\ 2.419e-01 & 9.522e-01 & -1.368e-01 & -1.271e-01 \\ 1.752e-01 & 4.531e-02 & 9.279e-01 & -3.259e-01 \\ 1.368e-02 & 1.667e-01 & 3.172e-01 & 9.335e-01 \end{bmatrix}.$$

CANONICAL ANALYSIS TABLE

	1	2	3	4
SLOPES	9.312	68.079	423.866	-285.804
CURVATURES	185.308	158.031	55.890	-78.763
AXIS DISTANCE	-0.025	-0.215	-3.792	-1.814
CHANGES IN y	186.008	197.175	736.281	501.254
FEATURE DISTANCE	4.209	4.204	1.827	3.798

APPENDIX 6.3j. The Second-Order Model of the Polysilicon/ Oxide Selectivity.

The second-order model was derived from a central composite experiment.

PARAMETER	DESCRIPTION	HIGH AXIAL VALUE	HIGH VALUE	CENTER VALUE	LOW VALUE	LOW AXIAL VALUE	UNIT OF MEASURE
A	Chlorine flow	57.1	55	50	45	42.9	sccm
B	Helium flow	57.1	55	50	45	42.9	sccm
C	Power	228.3	220	200	180	171.7	watts
D	Pressure	228.3	220	200	180	171.7	mTorr

The response unit of measure was Å polysilicon/Å oxide.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2	3	4
− 1.000	− 1.000	− 1.000	− 1.000	28.16	19.98	32.86	23.31
+1.000	− 1.000	− 1.000	− 1.000	25.63	25.66	25.49	25.53
− 1.000	+1.000	− 1.000	− 1.000	26.48	21.60	27.80	22.69
+1.000	+1.000	− 1.000	− 1.000	20.52	22.61	22.51	24.80
− 1.000	− 1.000	+1.000	− 1.000	19.63	19.79	19.94	20.10
+1.000	− 1.000	+1.000	− 1.000	18.44	18.33	20.01	19.89
− 1.000	+1.000	+1.000	− 1.000	17.02	18.88	14.68	16.28
+1.000	+1.000	+1.000	− 1.000	18.94	18.24	19.06	18.35
− 1.000	− 1.000	− 1.000	+1.000	33.60	27.83	35.38	29.31
+1.000	− 1.000	− 1.000	+1.000	28.47	32.02	29.43	33.10
− 1.000	+1.000	− 1.000	+1.000	26.32	30.32	26.35	30.36
+1.000	+1.000	− 1.000	+1.000	30.82	28.80	33.15	30.36
− 1.000	− 1.000	+1.000	+1.000	21.63	22.50	22.14	23.02
+1.000	− 1.000	+1.000	+1.000	25.40	23.69	27.23	25.40
− 1.000	+1.000	+1.000	+1.000	19.92	21.51	18.73	20.23
+1.000	+1.000	+1.000	+1.000	21.41	22.64	21.41	22.64
+1.414	0.000	0.025	0.050	34.37			
− 1.414	0.000	0.025	0.050	24.07			
0.000	+1.414	0.025	0.050	23.35			
0.000	− 1.414	0.025	0.050	22.73			
0.000	0.000	+1.391	0.050	28.14			
0.000	0.000	− 1.441	0.050	19.10			
0.000	0.000	0.025	+1.368	40.06			
0.000	0.000	0.025	− 1.469	30.11			
0.000	0.000	0.025	0.050	18.81			

The second-order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 29.85,$$

$$\mathbf{b}' = \begin{pmatrix} 3.308e-01 & -1.104e+00 & -3.730e+00 & 2.457e+00 \end{pmatrix},$$

$$\mathbf{B} = \begin{bmatrix} -2.529e+00 & 9.785e-02 & 2.225e-01 & 2.499e-01 \\ 9.785e-02 & -1.666e+00 & -5.481e-02 & -2.393e-02 \\ 2.225e-01 & -5.481e-02 & 4.018e-01 & -2.238e-01 \\ 2.499e-01 & -2.393e-02 & -2.238e-01 & -2.089e+00 \end{bmatrix}.$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION ß0	4.336e+04	1	4.336e+04	8.923e+03
R| ß0	1.634e+03	14	1.167e+02	2.403e+01
LACK OF FIT	1.182e+02	10	1.182e+01	2.432e+00
PURE ERROR	2.332e+02	48	4.859e+00	
TOTAL	4.534e+04	73		

STATIONARY POINT

x_0 = (4.544e-01 -4.529e-01 4.425e+00 1.735e-01).

CHARACTERISITIC ROOTS

λ = (4.363e-01 -1.655e+00 -1.982e+00 -2.681e+00).

M MATRIX

$$M = \begin{bmatrix} 6.704e-02 & 1.163e-01 & 4.214e-01 & 8.969e-01 \\ -2.188e-02 & 9.931e-01 & -5.437e-02 & -1.016e-02 \\ 9.942e-01 & 1.440e-02 & 4.432e-02 & -9.700e-02 \\ -8.129e-02 & 4.801e-03 & 9.042e-01 & -4.194e-01 \end{bmatrix}.$$

CANONICAL ANALYSIS TABLE

	1	2	3	4
SLOPES	-3.861	-1.100	2.255	-0.260
CURVATURES	0.436	-1.655	-1.982	-2.681
AXIS DISTANCE	4.425	-0.332	0.569	-0.048
CHANGES IN y	6.702	2.524	4.380	2.718
FEATURE DISTANCE	0.661	4.462	4.438	4.474

APPENDIX 6.3k. The Second-Order Model of the Polysilicon/ Photoresist Selectivity.

The second-order model was derived from a central composite experiment.

PARAMETER	DESCRIPTION	HIGH AXIAL VALUE	HIGH VALUE	CENTER VALUE	LOW VALUE	LOW AXIAL VALUE	UNIT OF MEASURE
A	Chlorine flow	57.1	55	50	45	42.9	sccm
B	Helium flow	57.1	55	50	45	42.9	sccm
C	Power	228.3	220	200	180	171.7	watts
D	Pressure	228.3	220	200	180	171.7	mTorr

The unit of measure of the response was Å polysilicon/Å photoresist.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2	3	4
− 1.000	− 1.000	− 1.000	− 1.000	0.91	0.96	1.04	1.04
+1.000	− 1.000	− 1.000	− 1.000	1.06	1.03	1.06	1.02
− 1.000	+1.000	− 1.000	− 1.000	1.01	0.94	1.06	0.95
+1.000	+1.000	− 1.000	− 1.000	1.10	1.01	1.21	1.10
− 1.000	− 1.000	+1.000	− 1.000	0.98	0.95	1.00	0.97
+1.000	− 1.000	+1.000	− 1.000	1.02	0.95	1.10	1.03
− 1.000	+1.000	+1.000	− 1.000	0.99	1.00	0.86	0.86
+1.000	+1.000	+1.000	− 1.000	0.97	1.04	0.98	1.04
− 1.000	− 1.000	− 1.000	+1.000	1.09	0.99	1.15	1.05
+1.000	− 1.000	− 1.000	+1.000	1.32	1.18	1.37	1.22
− 1.000	+1.000	− 1.000	+1.000	1.10	1.03	1.10	1.05
+1.000	+1.000	− 1.000	+1.000	1.16	1.23	1.25	1.33
− 1.000	− 1.000	+1.000	+1.000	1.06	0.98	1.09	1.01
+1.000	− 1.000	+1.000	+1.000	1.00	0.92	1.08	0.99
− 1.000	+1.000	+1.000	+1.000	1.00	1.03	0.94	0.97
+1.000	+1.000	+1.000	+1.000	0.91	1.02	0.91	1.02
+1.414	0.000	0.025	0.050	1.09			
− 1.414	0.000	0.025	0.050	1.01			
0.000	+1.414	0.025	0.050	1.11			
0.000	− 1.414	0.025	0.050	0.97			
0.000	0.000	+1.391	0.050	1.03			
0.000	0.000	− 1.441	0.050	1.18			
0.000	0.000	0.025	+1.368	1.11			
0.000	0.000	0.025	− 1.469	1.16			
0.000	0.000	0.025	0.050	1.08			

The second-order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 1.107,$$

$$\mathbf{b}' = \begin{pmatrix} 3.789e-01 & -3.874e-03 & -5.368e-02 & 3.302e-02 \end{pmatrix},$$

$$\mathbf{B} = \begin{bmatrix} -3.318e-02 & 2.458e-03 & -1.462e-02 & 5.017e-04 \\ 2.458e-03 & -3.723e-02 & -5.984e-03 & -3.602e-03 \\ -1.462e-02 & -5.984e-03 & -7.126e-03 & -1.535e-02 \\ 5.017e-04 & -3.602e-03 & -1.535e-02 & 1.441e-02 \end{bmatrix}.$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	8.019e+01	1	8.019e+01	2.274e+04
R I β_0	5.218e-01	14	3.727e-02	1.057e+ 01
LACK OF F IT	6.642e-02	10	6.642e-03	1.883e+ 00
PURE ERROR	1.693e-01	48	3.527e-03	
TOTAL	8.095e+01	73		

STATIONARY POINT

$x_0 =$ (1.124e+01 4.583e-01 -1.261e-00 -2.413e+00).

CHARACTERISITIC ROOTS

$\lambda =$ (2.338e-02 -7.067e-03 -3.861e-02 -4.083e-02).

M MATRIX

$$M = \begin{bmatrix} 1.362e\text{-}01 & -4.197e\text{-}01 & -4.785e\text{-}01 & 7.591e\text{-}01 \\ 3.670e\text{-}02 & -2.368e\text{-}01 & 8.758e\text{-}01 & 4.205e\text{-}01 \\ -4.971e\text{-}01 & 7.268e\text{-}01 & -2.853e\text{-}02 & 4.731e\text{-}01 \\ 8.569e\text{-}01 & 4.893e\text{-}01 & 5.577e\text{-}01 & 1.519e\text{-}01 \end{bmatrix}.$$

CANONICAL ANALYSIS TABLE

	1	2	3	4
SLOPES	0.060	-0.038	-0.018	0.007
CURVATURES	0.023	-0.007	-0.039	0.041
AXIS DISTANCE	-1.286	-2.678	-0.235	0.083
CHANGES IN y	0.107	0.066	0.050	0.042
FEATURE DISTANCE	2.689	1.310	2.972	2.980

APPENDIX 6.3I. The Second-Order Model of the Anisotropy

The second-order model was derived from a central composite experiment.

PARAMETER	DESCRIPTION	HIGH AXIAL VALUE	HIGH VALUE	CENTER VALUE	LOW VALUE	LOW AXIAL VALUE	UNIT OF MEASURE
A	Chlorine flow	57.1	55	50	45	42.9	sccm
B	Helium flow	57.1	55	50	45	42.9	sccm
C	Power	228.3	220	200	180	171.7	watts
D	Pressure	228.3	220	200	180	171.7	mTorr

The unit of measure of anisotropy was degrees.

EXPERIMENT DATA

ENCODED PARAMETERS (A,B,C,D)				RESPONSE OF REPLICATE 1	2
− 1.000	− 1.000	− 1.000	− 1.000	56.18	66.73
+1.000	− 1.000	− 1.000	− 1.000	65.77	68.20
− 1.000	+1.000	− 1.000	− 1.000	73.30	71.94
+1.000	+1.000	− 1.000	− 1.000	66.73	71.94
− 1.000	− 1.000	+1.000	− 1.000	59.89	56.98
+1.000	− 1.000	+1.000	− 1.000	63.89	63.89
− 1.000	+1.000	+1.000	− 1.000	56.98	57.38
+1.000	+1.000	+1.000	− 1.000	64.36	62.98
− 1.000	− 1.000	− 1.000	+1.000	69.19	63.89
+1.000	− 1.000	− 1.000	+1.000	74.36	73.83
− 1.000	+1.000	− 1.000	+1.000	68.20	64.95
+1.000	+1.000	− 1.000	+1.000	74.36	75.96
− 1.000	− 1.000	+1.000	+1.000	64.83	71.94
+1.000	− 1.000	+1.000	+1.000	68.20	63.89
− 1.000	+1.000	+1.000	+1.000	62.08	56.58
+1.000	+1.000	+1.000	+1.000	55.39	60.75
+1.414	0.000	0.025	0.050	48.98	
− 1.414	0.000	0.025	0.050	52.40	
0.000	+1.414	0.025	0.050	61.63	
0.000	− 1.414	0.025	0.050	63.89	
0.000	0.000	+1.391	0.050	42.02	
0.000	0.000	− 1.441	0.050	63.89	
0.000	0.000	0.025	+1.368	43.33	
0.000	0.000	0.025	− 1.469	59.04	
0.000	0.000	0.025	0.050	57.38	

The second-order model was written as

$$y = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}.$$

The model coefficients determined by the least squares regression were

$$b_0 = 4.479e+01,$$

$$\mathbf{b}' = \begin{pmatrix} 1.353e+00 & -2.894e-01 & -4.046e+00 & 5.637e-01 \end{pmatrix},$$

$$\mathbf{B} = \begin{bmatrix} 2.398e+00 & -1.772e-01 & -3.177e-01 & -5.950e-02 \\ -1.772e-01 & 8.435e+00 & -1.036e+00 & -8.752e-01 \\ -3.177e-01 & -1.036e+00 & 3.505e+00 & -1.236e-01 \\ -5.950e-02 & -8.752e-01 & -1.236e-01 & 2.937e-00 \end{bmatrix}.$$

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION β_0	1.634e+05	1	1.634e+05	1.634e+05
R | β_0	1.831e+ 03	14	1.308e+ 02	1.283e+ 01
LACK OF FIT	5.134e+ 02	10	5.134e+ 01	5.034e+00
PURE ERROR	1.632e+ 02	16	1.020e+ 01	
TOTAL	1.659e+05	41		

STATIONARY POINT

$x_0 =$ (-2.007e-01 7.881e-02 5.806e-01 -5.211e-02).

CHARACTERISITIC ROOTS

$\lambda =$ (8.765e+00 3.508e+00 2.774e+00 2.228e+00).

M MATRIX

$$M = \begin{bmatrix} -1.639e-02 & -2.562e-01 & -3.919e-01 & 8.835e-01 \\ 9.719e-01 & 1.172e-01 & 1.621e-01 & 1.239e-01 \\ -1.870e-01 & 8.946e-01 & 2.083e-01 & 3.484e-01 \\ -1.418e-01 & -3.470e-01 & -8.813e-01 & 2.878e-01 \end{bmatrix}.$$

CANONICAL ANALYSIS TABLE

	1	2	3	4
SLOPES	0.373	-4.196	-0.923	-0.088
CURVATURES	8.765	3.508	2.774	2.228
AXIS DISTANCE	-0.021	0.598	0.166	0.020
CHANGES IN y	8.788	8.070	3.202	2.233
FEATURE DISTANCE	0.621	0.169	0.599	0.621

Appendix 6.4. Anisotropy Adjustments.

In the course of conducting the etching experiments, it was determined that the anisotropy was highly dependent on the initial profile of the photoresist. This could be concluded directly from the fact that the polysilicon/photoresist selectivity did not vary much from unity. As assumptions, if we considered the etch mechanisms of both the photoresist and the polysilicon to be perfectly anisotropic, then the final profile of the polysilicon layer after etching would depend solely on two factors, the initial profile of the photoresist and the selectivity of the etch. These assumptions are not far from reality. The profiles of all the etches considered in the study seemed to demonstrate this dependence. In no observed cross section was there any evidence of polysilicon being etched beneath the photoresist, and the relative difference in the polysilicon and the photoresist profiles followed the variation in selectivity. (See Figure 6.6.4.1.)

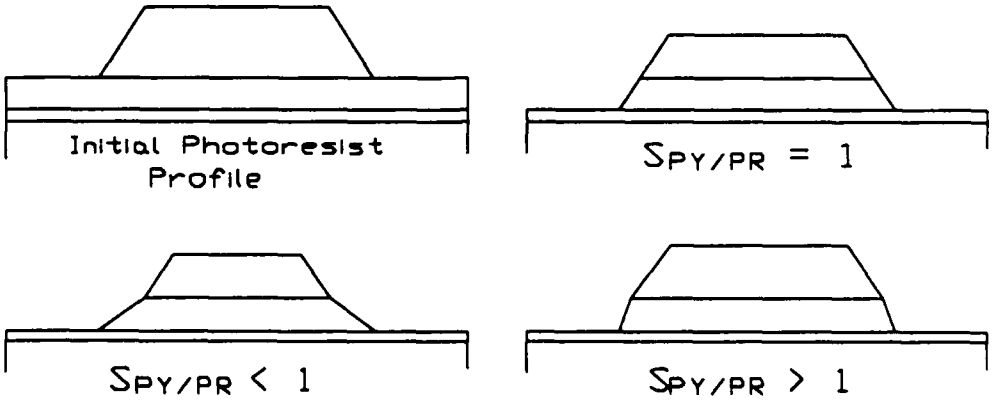


Figure 6.6.4.1 The Effect of PY/PR Selectivity on Anisotropy.

The observation of the etch dependence on the initial profile of the photoresist was used as the justification to adjust the measured responses of the anisotropy of the center and the axial experiments. The initial profile of the photoresist in these latter experiments was very different from the profiles of the photoresist in the initial full factorial experiments. As in the full factorial experiments, one of the wafers processed through the lithography with this latter batch was analyzed for this initial condition.[1] The angle of the photoresist profile for this wafer was 52°, 10° less than that of the wafers used in the full factorial experiments. This difference in the initial profile angle was a biasing element that made the data of the latter experiments unusable in the least squares regression. Since redoing the experiments would have been expensive in both time and materials, a second option of adjusting the anisotropy measurements was attempted.

The anisotropy adjustment was based on the assumption that the etch mechanisms of the photoresist and polysilicon etches were perfectly anisotropic. This is a reasonable assumption when the lateral etch of the polysilicon is dominated by the photoresist's erosion and when the photoresist's erosion is significantly greater in the vertical direction than in the horizontal direction. For any given etch, this assumption becomes more true as the initial photoresist angle decreases. If this condition were true, the anisotropy would be a function of the selectivity of the polysilicon etch over the photoresist etch and the initial profile angle of the photoresist. From this

1 This analysis was formulated after all the center and axial experiments had been completed, in an attempt to understand why the anisotropy of these experiments was so poor.

observation, it is apparent that one can calculate the selectivity of an etch from a measured anisotropy and an initial profile angle. Consider the anisotropy measurement described in Chapter 5:

$$A = (1- (run/rise)) \ .$$

Now consider what the expected anisotropy would be if the etch selectivity were 1, $S_{PY/PR}=1$. The anisotropy could be determined from the equation

$$A = (1 - \cot \theta)$$

where θ is the initial angle of the photoresist profile. The difference of these two measurements would be a measure of the anisotropy as seen in the variation of the lateral etch from the expected lateral etch.

$$(1 - (run/rise))-(1 - \cot \theta) = (\text{expected run - run})/\text{rise} \ ^*$$

$$= \Delta x/\text{rise}.$$

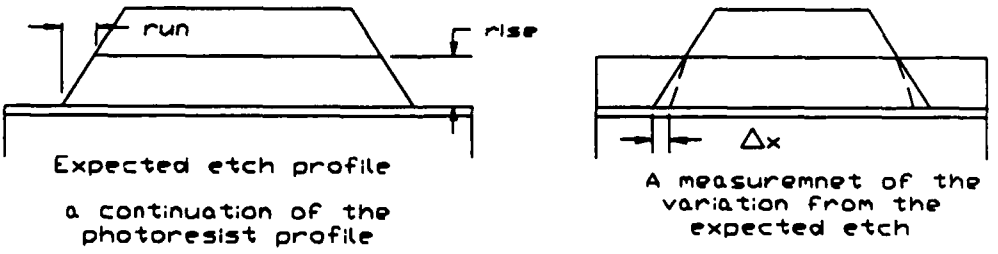


Figure 6.6.4.2 The Measurements Used to Analyze Anisotropy.

It is possible to obtain this same relation from the relative etch rates and the initial photoresist angle. Consider y to be the photoresist etch rate and z to

be the polysilicon etch rate. Then x, the rate of change in the lateral etch from the expected, is

$$x = \cot\theta\,(y-z)$$

The total difference can be determined from the length of time the wafer is exposed to the etch. In these experiments, this time was a function of the etch rate and the polysilicon thickness

$$t = \text{rise}/z,$$

so

$$\Delta x = xt = \cot\theta\,(y-z)\,(\text{rise}/z).$$

From this equation we can arrive at an equivalent expression for $\Delta x/\text{rise}$,

$$\Delta x/\text{rise} = \cot\theta\,((y-z)/z) \quad **$$

Setting the two equations * and ** equal to each other we have

$$((y-z)/z) = (A - (1-\cot\theta))/\cot\theta.$$

The expression on the left side of the equation is considered a constant regardless of the initial photoresist profile angle, and can, therefore, be used as a constant value to compare two etches with different initial conditions.

$$\frac{A_1 - (1 - \cot\theta_1)}{\cot\theta_1} = \frac{A_2 - (1 - \cot\theta_2)}{\cot\theta_2}$$

$$A_1 = \frac{(A_2 - (1 - \cot \theta_2))(\cot \theta_1) + (1 - \cot \theta_1)}{\cot \theta_2}$$

So, the equation for adjusting the anisotropy measurement from a value when the initial photoresist profile was 52° to an expected value if the initial profile was 62° is

$$A_1 = (((A_2 - .23).50)/.77) + .50,$$

or

$$A_1 = .65\ A_2 + .35.$$

CHAPTER 7

CONCLUSION

Although a definitive point has not been identified as the optimum operating point for the reactive ion etching of polysilicon in a chlorine and helium plasma, the process has been well studied and an optimizing strategy, response surface methodology, has been explained and applied to the process. From the research that has been completed, much has been learned about the process and the factors that influence the process results. Additionally, the research has confirmed the fact that this etching process has the potential to achieve the characteristics that are desired, namely high anisotropy and high polysilicon-to-oxide selectivity. Finally, a recommended course of action has been suggested for continuing the optimizing study. RSM has not only been useful in determining the effects of the different parameters, it has also been demonstrated to be a framework through which one can search for optimum conditions. But the usefulness of response surface methodology does not stop here. It also has the potential to provide a process model that can be directly inserted into adaptive control algorithms for sustained process control and improvement.

PARAMETER EFFECTS

The results of the research confirm the assumption that the four parameters; chlorine flow rate, helium flow rate, power, and pressure, were significant in their effects on the process. Each had a role to play in the process. The results also confirm that there were significant interactions

amongst the parameters of the process.[1] But the results reveal that the effects of these four parameters and their interactions were not the only significant factors that affected the process. The initial photoresist profile had a dominant effect on the anisotropy of the etch. The observed effects of the parameters are summarized below.

CHLORINE FLOW RATE

The effects of the chlorine flow rate were surprising. All etch rates decreased with the increase in chlorine. The response that was most affected by the chlorine was the etch of the photoresist, not the etch of the polysilicon. These observations have some significant implications. The interaction of the chlorine with the photoresist tends to provide some protection against etching. It not only occurs on the photoresist surface but on the polysilicon and oxide surfaces as well. Chlorine, its ions, and its radicals are the only chemical species that react with any of the layers to cause etching. One would suspect that an increase in chlorine would increase the polysilicon etch rate. The fact that it does not supports the hypothesis that the chlorine-photoresist interaction creates a polymer layer on the wafer surface that protects against etching. Etching did not stop completely since sputtering mechanisms were still present. However, if regions of the wafer were protected from the sputtering mechanisms, such as the polysilicon sidewalls, then there would be no etching. Although not verifiable in this experiment due to the poor initial photoresist profiles, one can conclude that

1 Appendices 6.1a-f are the results of the evaluation of parameter significance using the full factorial design.

the mere presence of photoresist in this etching process enhances the anisotropy of the polysilicon etch.

HELIUM FLOW RATE

The main effects of the helium flow rate were the increases of all the etch rates. Since the oxide etch rate increased by a greater percentage than the polysilicon etch rate, the polysilicon/oxide selectivity decreased with the increased helium flow rate. But since the change in the photoresist etch rate was not much different than the change in the polysilicon etch rate, the polysilicon/photoresist selectivity was not sensitive to the increase in the helium flow rate. The results also show that the anisotropy was not sensitive to the helium flow rate. (This may have been the result of the dominance of the initial photoresist profile anisotropy.)

These results confirm that the helium flow rate most affects the physical mechanisms of the etch. Helium, an inert gas, does not chemically react with the layers of the wafer. It can, however, influence the chemical reactions in its influence on the formation of the plasma. But since the oxide etch rate is most increased and since the oxide is least sensitive to the chemicals in the etch, the physical mechanisms are shown to be most affected.

POWER

Of the four parameters evaluated, power is by far the most influential. It significantly affected all of the responses that were evaluated. As expected, increased power increased all etch rates. But since the power's effect on the photoresist and the oxide etch rates was greater than its effect on the

polysilicon etch rate, all the selectivities and the anisotropy decreased with the increase in power. (Again, the effect on anisotropy may have been the result of the initial photoresist profile.)

These results imply that the power most affected the physical mechanisms of the etch. They further imply that the optimum etch that was sought would be achieved at a lower power than that currently used in the etch recipe.

PRESSURE

The main effects of the increase in pressure were the decreases in all of the etch rates. Since the decrease in the photoresist and the oxide etch rates were greater than the decrease in the polysilicon etch rate, the selectivities and anisotropy increased with higher pressure.

These results help confirm that increased pressure reduces the physical mechanisms in the etch and increase the absorption of chlorine at the wafer surface. Both of these mechanisms would decrease the etch rate of the photoresist and indeed the photoresist etch rate decreased sharply. These two mechanisms would counteract each other at the polysilicon surface. This too is confirmed by the results. The polysilicon etch rate is least affected by the pressure while the oxide etch rate, which would only be affected by the decrease in physical mechanisms, decreased rapidly.

Table 7.1 is a rewritten version of Table 1.2 with the experimentally obtained effects of the four evaluated parameters.

Table 7.1 The Observed Parameter Effects.

	Polysilicon Etch Rate	Oxide Etch Rate	Photoresist Etch Rate	Polysilicon/ Oxide Selectivity	Polysilicon/ Photoresist Selectivity	Anisotropy
HELIUM	+	+	+	–	0	0
CHLORINE	–	–	–	+	+	+
POWER	+	+	+	–	–	–
PRESSURE	–	–	–	+	+	+

PARAMETER INTERACTIONS

The results of these experiments also confirm the presence of interactions among the parameters. If one simply considered the results summarized in Table 7.1, it would seem an optimum etch could be obtained by simply decreasing the helium flow rate and the pressure. But the results show there is a delicate balance that must occur between the physical and chemical mechanisms. Since all four parameters affect both of these mechanisms there is also a delicate balance between them.

The application of RSM to the study of this process helped reveal the balance of the parameters that was needed. Unfortunately, the study stopped short of finding an optimum set of conditions. But the study did reveal a region where an optimum etch may be found. (See Chapter 6.) In order to find the optimum point, however, more experiments must be conducted. But before any more etching experiments are conducted the issue of the initial

photoresist profile must be resolved. Additionally, the actual process for which the etch is to be used must be considered.

PHOTORESIST PROFILE

As explained in detail in Chapter 6, the initial photoresist profile significantly influenced the anisotropy of the polysilicon etch. As the initial photoresist profile became more vertical the polysilicon etch became more anisotropic. Since the photoresist profiles were far from vertical, these profiles were the dominant factors influencing the etch anisotropy; the implication being that to optimize the etch anisotropy one must first optimize the photolithography. As a recommended place to start, one should consider the illustrations of [6:453] which show that a considerable reflow of the photoresist, Shipley 1350-J, occurs at the 125°C post bake temperature that was used in the photolithography process. A lower post bake temperature may improve the results.

OTHER CONSIDERATIONS

Other factors must be considered in the continued study of this etching process. The initial conditions of the etch in a real application are most likely to be different than those used in this study. In this study, undoped polysilicon was etched. In a real application, the polysilicon would be doped. This has two implications. First of all, n doped polysilicon will usually etch faster than undoped polysilicon in any given etch. The balance of the vertical and the lateral etching mechanisms of the polysilicon may be different. Second, if the dopant is diffused in an oxidizing ambient, there will be an oxide grown on the polysilicon surface. This oxide has potential as a

mask in the etch. The use of an oxide mask would make the etch less sensitive to the initial photoresist profiles, but the etching sequence would have to be revised such that the oxide could be etched. Either an additional etching process would have to be introduced or the CF_4 etch would have to be lengthened in order to etch the oxide. These changes would not be significant since they could all be added to the etching sequence. The plasma Technology RIE 80 M can control up to three consecutive processes.

.

THE VALUE OF RSM

Regardless of the initial conditions of the wafer prior to the polysilicon etching, the best way to find the optimum etch is to use RSM. Since there is no available physical model of the reactive ion etching of polysilicon using a chlorine and helium plasma, an empirical approach must be used to optimize the process. RSM provides a procedure to understand and optimize processes empirically.

THE FURTHER APPLICATION OF RESPONSE SURFACE METHODOLOGY

At the completion of a response surface study the experimenter identifies a set of parameter settings which he believes should be used in running the evaluated process to obtain the optimum responses. Often these parameter settings are chosen based on a best understanding of the process. Reasons such as the cost of experimentation or the need to get production started may cause the experimenter to choose settings which may be short of the optimum. Maybe the optimized process was developed

in laboratory conditions and then applied to larger manufacturing equipment where the process becomes slightly different. Or, perhaps, over time, wear on the equipment or the buildup of residues in the equipment cause slow and continuous changes in the process. From these situations the question arises, "Can we continue the response surface study while the process is being used in production?"

One method to continue the response surface study was suggested by Box in [7] and is called Evolutionary Operations (EVOP). In EVOP the response surface study continues on the production line where perturbations to the controlled parameters are kept sufficiently small such that the product remains within specifications. In other words the process is continuously evaluated with RSM but the parameter space is kept very small so that the responses do not change very much. Several problems exist in executing this approach. If the measurement or the process errors are large compared to allowed changes in the responses, a usable model may not be achievable. If larger changes in the responses are allowed, it may be difficult to ensure that the product remains within specifications. However, the biggest drawback of this approach is that it prevents the application of statistical process control. Changes in the process due to uncontrolled parameters cannot be distinguished from the changes caused by the controlled parameter variations in the designed experiments. As can be seen, EVOP is very intrusive; it continuously affects the process as it is applied.

A better approach from the production point of view would be to use run data to continuously improve the fitted models. If this is done, statistical process control can be applied since the study of the process is not intrusive.

The study of the process would not mandate the variation of the controlled parameters, and data on the distribution of the errors of the responses could be maintained. A proposed method for applying this approach is to periodically refit the model using the set point of the runs as a new data point and the responses as the additional data. The new model could then be reanalyzed for optimum conditions using the methods discussed in this thesis. Of course, the process would have to be taken off line for a few runs so that verification experiments could be conducted. But, this exploration would be short and production could be restarted with a theoretically better process. Some disadvantages of this approach are the eventual loss of rotatability of the model and the eventual dominance of the data in the vicinity of the runs. The consequences of these factors have not been fully explored and offer an area for further study.

Another proposed approach which is very similar to the second above borrows a technique used in adaptive control called recursive least squares (RLS). The theory of RLS is well presented by Karl Åström and Björn Wittenmark in [8:60-69]. This method differs from the second approach in that the model is refitted using RLS rather than the complete least squares regression. In RLS the average of the responses of an interval of runs and the set point is used in an algorithm that revises the model coefficients. The RLS algorithm is based on the same general model that is used in least squares regression

$$y = X\beta + \varepsilon,$$

but the calculations are done in a recursive manner where the coefficients in the model are continuously updated. This equation, therefore, is rewritten to account for recursion as

$$y(n) = X(n)\ \beta(n) + \varepsilon(n).$$

By defining

$$P(n) = (X'(n)X(n))^{-1},$$

the equation for determining the coefficients that minimize the error becomes

$$\hat{\beta}(n) = P(n)X'(n)\ y(n). \quad (1)$$

In RLS the recursion occurs in the updating of P(n) with each experiment. But unlike the second approach discussed previously, the P(n) is not updated by *adding a datapoint to the X matrix. P(n) is updated by the equation*

$$P(n)^{-1} = P(n-1)^{-1} + q(n)q'(n) \quad (2),$$

where q(n) is a vector formed in the same manner as the rows of the X matrix using the parameter values of the n^{th} interval. (1) then becomes

$$\hat{\beta}(n) = P(n)(\sum_{i=1}^{n-1} q(i)y(i) + q(n)y(n)). \quad (3)$$

The expression

$$\sum_{i=1}^{n-1} q(i)y(i)$$

is equivalent to

$$X'(n)y(n-1),$$

so from (1) and (2) it follows that

$$X'(n-1)y(n-1) = P(n-1)^{-1}\,\hat{\beta}(n-1) = P(n)\,\hat{\beta}(n-1) - q(n)q'(n)\,\hat{\beta}(n-1).$$

Equation (3) then becomes

$$\begin{aligned}\hat{\beta}(n) &= \hat{\beta}(n-1) - P(n)q(n)q'(n)\,\hat{\beta}(n-1) + P(n)q(n)y(n)\\ &= \hat{\beta}(n-1) + P(n)q(n)(y(n) - q'(n)\,\hat{\beta}(n-1)).\end{aligned}$$

Note that

$$(y(n) - q'(n)\hat{\beta}(n-1)) = \varepsilon(n).$$

So the new model coefficients are a function of the original model coefficients, the set point of the interval considered, and the residual at the interval set point.

This latter approach is a fertile area for research. Unlike the second approach it does not revise the model coefficients for a global view but attempts to revise the model such that the model predictions and the actual responses converge at the region in the parameter space where the process is run. Such an approach offers an adaptive model that will account for the variations in the process due to the equipment wear and the buildup of residues but may not be very effective in generating a model that is better from the perspective of response surface analysis. This approach may be more effective when the operating point is the extreme of a ridge or valley as

opposed to a point on a ridge or valley. And finally, in the initial applications of this procedure, the response may jump around too much to be of use. Of course, none of these points can be stated with certainty. The application of RLS deserves additional study.

CONCLUSION

In this thesis RSM has been shown to be an effective technique for optimizing processes empirically. It has been applied to a specific process, the reactive ion etching of polysilicon. From this study, the etching process is better understood and a region for improved responses has been identified. But the use of RSM does not need to stop at this point. RSM offers a methodology for continued improvement of processes and it may even have the potential to assist in the improvement and control of processes as they are used in production.

REFERENCES

[1] George E.P. Box, and Norman R. Draper. Emperical Model Building and Response Surfaces. New York: John Wiley & Sons, 1987.

[2] George E. P. Box, William G. Hunter, and J Stuart Hunter. Statistics for Experimenters. New York: John Wiley & Sons, 1978.

[3] Russ A. Morgan. Plasma Etching in Semiconductor Fabrication. New York: Elsevier Science Publishing Company, Inc., 1985.

[4] Douglas C. Montgomery. Design and Analysis of Experiments. New York: John Wiley & Sons, 1984.

[5] Raymond Meyers. Response Surface Methodology. Boston: Allyn and Bacon, Inc. 1971.

[6] Stanley Wolf and Richard Tauber, Silicon Processing for the VLSI Era. Sunset Beach, CA: Lattice Press, 1987.

[7] George E. P. Box, "Evolutionary Operation: A Method for Increasing Industrial Productivity," in The Collected Works of George E. P. Box. ed. George C. Tiao. Belmont, CA: Wadsworth, Inc., 1985. II,649-669.

[8] Karl J. Åström and Björn Wittenmark, Adaptive Control. New York: Addison-Wesley Publishing Company, 1989.

REFERENCES NOT CITED

[9] William E. Biles and James J. Swain. Optimization and Industrial Experimentation. New York: John Wiley & Sons, 1980.

[10] George E. P. Box, The Collected Works of George E. P. Box. 2 vols, ed. George C. Tiao. Belmont, CA: Wadsworth, Inc., 1985.

[11] Jenkins et al. "The Modeling of Plasma Etching Processes Using Response Surface Methodology," Solid State Technology. (April 1986), 175-181.

VITA

John Andrew Stine [redacted], the son of Mary Biggers Stine and Carl William Stine. After completing his work at Christopher Columbus High School, Miami, Florida, in 1977, he entered the United States Military Academy at West Point, NY. In May, 1981, he received the degree of Bachelor of Science in Engineering from West Point and was commissioned as a Second Lieutenant in the United States Army Corps of Engineers. He served overseas in the Third Infantry (Mechanized) Division in West Germany from 1982 to 1985 and then served stateside in the Second Armored Division at Fort Hood, Texas, until June, 1988. He commanded C Company, 17th Engineer Battalion, 2AD from June, 1986 to December, 1987. In April 1985, he married the former Martha Jane Byrd and today they have three sons Andrew, Harry, and Brian. In June, 1988, he entered the graduate school of The University Of Texas. In May, 1990, he received the degree of Master of Science in Engineering with the major of Manufacturing Systems Engineering.

[redacted]

This thesis was typed by John A. Stine.